

Vermont Million Hearts Report

Examining the Utility of Vermont's All Payer Claims Database

October 2014

Supported by a grant from the Association of State and Territorial Health Officials



Executive Summary

The Vermont Healthcare Claims Uniform Reporting and Evaluation System (VHCURES) is a new data system for public health analysts at the Vermont Department of health. VHCURES includes only paid claims that most insurers are required to report to a state database. All private insurers that cover more than 200 Vermont lives are required to report to the database. Medicaid paid claims also appear in VHCURES and Medicare paid claims are expected to be included in the near future. Other federal insurance plans such as TRICARE do not appear in VHCURES.

VHCURES became available to Department of Health analysts in November of 2012. While there is a user's group for all state employees and contractors who utilize VHCURES data, it is one of the more complex data systems available at the Health Department. Adding to the data structure complexities is that many of the groups using VHCURES have taken additional steps to create their own versions of the data for use on analytic projects. These steps can ultimately affect how data is included for analysis and which variables are used for various data procedures. While this is a normal part of a data analysis, these additional data manipulations are not always transparent to VHCURES data users.

Some analysts at the Health Department have performed basic analyses using VHCURES data, but there continue to be questions. Based on some of the experiences of the health department analysts in the Division of Health Surveillance and the data needs of the Division of Health Promotion and Disease Prevention related to VHCURES the following goals were developed at the start of the Million Hearts acceleration project in February 2014:

1. Improve the surveillance of hypertension, heart attack, and stroke;
2. Better inform public health priorities and activities;
3. Assist clinical partners in care management and quality improvement of hypertension control

To accomplish these goals, the Department of Health and Truven Health Analytics (Truven), a contractor familiar with VHCURES analysis, performed two case studies and compiled a document describing the potential for linking various data sources.

The first case study involved collaboration between the Department of Health, Truven, and the Community Health Centers of the Rutland Region to compare electronic health record data to claims data during a similar time period. All three stakeholders were involved in developing an analysis plan for how the electronic health record (EHR) and the VHCURES data would be structured for comparison. Truven was responsible for compiling the final version of the analysis plan and the final report. The Department of Health was responsible for learning and replicating the VHCURES data and making sure the EHR data was analyzed according to the analysis plan. The Health Department also over saw the drafting and finalizing of all reports produced by Truven.

The second case study focused on comparing Hospital discharge data from the Vermont Uniform Hospital Discharge Data set (VUHDDS) to VHCURES. Similar to case study one, the Department of Health and Truven collaborated to make the analysis plan. The Health Department Analyst performed all analyses on the hospital data and had regular discussions to learn specifically how data was run in

VHCURES. Due to unintended results, more discussion was needed to determine the overall results and compile the final report for this case study 2.

Lastly, the document that described the potential for linking data sources, referred to as the “Linkage Memo”, involved both the Health Department Analyst and Contractor staff discussion. The Department of Health Analyst discussed resources for learning about data sources via the health department website and also shared the Data Encyclopedia, a document that describes commonly accessed data sources in the Division of Health Surveillance.

At the end of the project the Analyst presented the project findings to an audience that included both program managers and other analysts who were both familiar and unfamiliar with VHCURES data. The presentation gave a high level overview of both case studies and the “Linkage Memo”. After presenting this information it became apparent that more clarification on linking versus comparing data was needed and some additional documentation on the topic was written by the Health Statistics Chief.

While each case study, the “Linkage Memo”, and the additional description can be considered independent documents, they are better combined together, giving a broader picture of the role of VHCURES. For this reason, we have compiled them into the attached information packet for distribution throughout the Health Department and to external state, and national partners.

Vermont Million Hearts Report

Examining the Utility of Vermont's All Payer Claims Data base

Table of Contents

Case Study 1	5
Case Study 2	15
Linkage Memo	25
<i>Appendix A : List of Acronyms</i>	32
<i>Appendix B: Health Surveillance Data Encyclopedia</i>	33
Memo for Health Statistics Chief RE: Definition of Data Linkages	69
Using Vermont's All-Payer Claims Database: A Training for Public Health Professionals	73
Lessons Learned--VHCURES utility for chronic disease	77

Vermont Million Hearts Report

Examining the Utility of Vermont's All Payer Claims Data base

Vermont Million Hearts: Case Study #1 Report

Vermont Million Hearts: Case Study #1 Report

Objective

The purpose of this case study is to generate and compare population health care statistics obtained from two distinct data sources: the Vermont Healthcare Claims Uniform Reporting Evaluation System (VHCURES) and the Rutland Vermont Electronic Health Records (EHR) system. Specifically, we calculate the prevalence of hypertension and diabetes among the population served by the Rutland Federally Qualified Health Center (FQHC) in 2012 using each data source and compare findings.

VHCURES and the Rutland EHR system can each characterize the health status of Rutland area residents. Quantifying the differences in findings by data source provides insight into how the two sources may complement each other to provide Vermont policymakers with a more complete picture of health status than either source can offer individually.

Data Sources

VHCURES, Vermont's all-payer claims database, includes health insurance enrollment and paid claims data for Vermont residents covered by insurance companies with more than 200 enrollees in Vermont, excluding Tri-care, military, Veteran's Administration and the uninsured. For this case study, we used the December 2013 release of VHCURES released by Onpoint Data Systems, Portland, ME. This version is limited to enrollment and claims data from Medicaid and commercial carriers.

The Rutland EHR system contains health information for all patients who have visited the Rutland FQHC, regardless of insurance coverage and residence. This EHR system, located on site at the Rutland FQHC, stores detailed patient medical histories, including all diagnoses assessed and procedures performed at the Rutland FQHC or reported by the patient.

There are several notable differences between the two data sources that may influence prevalence rates. First, VHCURES contains enrollment data for all individuals covered by payers that meet the APCD reporting criteria, which means that VHCURES can account for those commercial plan and Medicaid enrollees that never used the health care system. In contrast, the Rutland EHR data only includes individuals who have received care from the Rutland FQHC, regardless of insurance status, including the uninsured and those covered by Medicare. Second, VHCURES contains claims information about care provided in different settings (e.g., clinic, hospital, lab) and from different providers. By contrast, the Rutland EHR data only includes information on care provided at the FQHC itself, although there may be some instances where Rutland FQHC patients have reported care provided by other providers that were recorded in the EHR. Third, the VHCURES data is based on billing data provided by the provider to the insurance company for reimbursement purposes. Claims include standard diagnostic and procedure codes as determined by medical coders based on clinical data. In contrast, the EHR data is rich clinical data recorded by the patient's care team that includes patient history, lab results, diagnoses, procedures provided or recommended, and any other clinically important information collected during the patient's visit.

Inclusion Criteria

In order to generate comparable statistics from VHCURES and the Rutland EHR given the differences described in the last section, and because we did not have access to the VHCURES Medicare data, we

restricted the study population to Vermont residents between ages 25 and 64 as of 2012 who lived in a zip code served by the Rutland FQHC catchment area (Appendix A) and were enrolled in either a commercial health insurance plan or Medicaid in 2012.

In the VHCURES analysis, we limited the denominator to only those enrollees who had at least one eligible claim in 2012. Eligible claims are defined as claims associated with either a facility-based outpatient clinic or a physician's office paid using either a commercial plan or Medicaid as the primary payer. In the EHR analysis, we only included individuals that had at least one office visit to the Rutland FQHC in 2012.

Hypertension Results

Examination of the hypertension results from the two data sources (Table 1) shows that the VHCURES population is slightly larger than the Rutland EHR population (21,356 VHCURES vs. 20,248 EHR). This was expected because the VHCURES population includes all enrollees residing in the Rutland service area with at least one outpatient claim whereas the Rutland EHR data includes only those Rutland service area residents who had visited the Rutland FQHC. These greater population counts in VHCURES relative to the Rutland EHR were present across all patient characteristic (age, gender, payer) with two exceptions: 25 to 34 year olds (4,560 VHCURES vs. 5,014 EHR) and men (9,284 VHCURES vs. 9,785 EHR). This could be because these are two population groups that tend to have the high rates of uninsurance. More specifically, if Rutland FQHC patients in either group lacked coverage at the time of their 2012 visit but have since enrolled in either a commercial plan or Medicaid, they would have been included in the EHR study population. Because they were uninsured in 2012, however, they would not have been included in the VHCURES analysis.

The VHCURES data had higher counts of individuals with hypertension, both overall and among each stratum, relative to the Rutland EHR numbers. As with the larger population counts, these higher counts of individuals with hypertension observed in the VHCURES data may be attributable to the fact that the VHCURES population includes individuals that visited providers other than the Rutland FQHC.

The overall prevalence of hypertension calculated from VHCURES (22%) was slightly higher than that using EHR (20%). The VHCURES data had higher hypertension prevalence estimates compared to the Rutland EHR system among all strata except for females (for which hypertension rates were nearly equal). All prevalence differences between the two sources other than for males were within 2 percentage points of each other. The disparity in estimated rates for males with hypertension was 5 percentage points—the largest gap observed. One potential explanation for this finding is that Rutland area residents who received care outside of the Rutland FQHC had a higher prevalence of hypertension as compared to Rutland area residents that received care at the Rutland FQHC.

Diabetes Results

Table 2 displays diabetes estimates calculated from VHCURES and the Rutland EHR system. The diabetes prevalence was slightly higher from VHCURES (9.0%) compared to the Rutland EHR system (7.4%) As with the hypertension tabulations, VHCURES prevalence estimates were a bit higher across most strata—by 2 percentage points on average.

Table 3 shows the proportion of individuals with at least one HbA1c test that had diabetes and the proportion of individuals with at least one HbA1c test that did not have diabetes. The HbA1c test is used

to determine the average amount of sugar in a patient's blood by measuring the patient's hemoglobin A1C level. It is traditionally given to patients with diabetes in order to monitor how well their diabetes is being controlled. Increasingly, however, HbA1c testing is also being used to screen patients for diabetes.

The proportion of individuals that received at least one HbA1c test in 2012 is estimated to be 12% using VHCURES and 10% using the Rutland EHR data. However, estimates of diabetes prevalence among individuals who received at least one HbA1c test in 2012 differed between the two sources. The VHCURES HbA1c data showed that 38% of the people who received the test did not have diabetes and 62% did have diabetes. By contrast, the Rutland EHR HbA1c data showed that 52% did not have diabetes and 48% had diabetes.

This difference may be related to the use of HbA1c testing as a screening procedure. In these cases, diabetes would not be recorded on the problem list. However, if the HbA1c test is performed as a "rule out" procedure, a diabetes diagnosis code may have been captured in the insurance claim. The disparity between source estimates narrows with age. This is consistent with this explanation as diabetes prevalence increases with age, and hence the difference between VHCURES and EHR prevalence calculations would diminish with age. We continue to explore other potential explanations for this anomaly between data sources.

Conclusion

VHCURES, Vermont's all-payer claims database, and the Rutland FQHC EHR system yielded very similar prevalence estimates for diabetes and hypertension when the same inclusion criteria were applied to both data sources. The VHCURES prevalence estimates tend to be slightly greater across the board, except for HbA1c testing. Given the fact that administrative claims often do not have the depth of information that medical records contain, more study is needed to see if that finding is an anomaly.

In our analysis, we used two data sources, each with its own limitations. VHCURES, like all administrative data, contains less detail, is more prone to miscoding, and includes only those individuals covered by a subset of insurance plans. Electronic medical records, like the one used by the Rutland EHR, contain more detail on a patient but are usually limited to the health care experience of individuals who visit the particular facility with which the system is associated and may provide a less-than-complete picture of population health. Additionally, there is no single EHR system: in this case, key information such as diagnosis and enrollment date limited the ability to perform longitudinal analyses.

Our comparison of VHCURES and the Rutland EHR showed how one could use the findings from two limited databases as validation for one another: for VHCURES, this study represented an audit of the accuracy of the estimates generated by the claims using the "gold standard" of medical records; for the Rutland EHR, the comparison to VHCURES enabled assessment of how well a single clinic was doing in capturing population health. In addition, the clinical information available from the EHR data can be helpful for understanding the clinical needs of specific populations identified. For example, the EHR data could provide the HbA1c values of the population that was screened for this test but did not have a diabetes diagnosis. In addition, the expenditures information available from claims data can be helpful for understanding the financial impact of changes to treatment protocols, such as those that lower the threshold for prescribing drugs.

While researchers must be vigilant when assessing statistics produced from either type of data, these very similar findings indicate that this technique of comparing two complementary databases could be a useful and robust tool to use when developing or validating new measures of health in the future.

Table 1: Population Counts and Hypertension

	Population with at least one outpatient claim			Individuals with hypertension			Prevalence of hypertension based on having at least one outpatient claim		
	VHCURES	Rutland EHR System	<i>Difference</i>	VHCURES	Rutland EHR System	<i>Difference</i>	VHCURES	Rutland EHR System	<i>Difference</i>
TOTAL	21,356	20,248	1,108	4,687	4,055	632	21.95%	20.00%	1.95%
Age Group									
25-34	4,491	5,014	-523	227	192	35	5.05%	3.83%	1.22%
35-44	4,560	4,193	367	600	545	55	13.16%	13.00%	0.16%
45-54	6,274	5,599	675	1,523	1,298	225	24.27%	23.18%	1.09%
55-65	6,031	5,442	589	2,337	2,020	317	38.75%	37.12%	1.63%
Gender									
Female	12,072	10,463	1,609	2,144	1,864	280	17.76%	17.82%	-0.06%
Male	9,284	9,785	-501	2,543	2,191	352	27.39%	22.39%	5.00%
Payer									
Commercial	14,897	-	-	3,627	-	-	24.35%	-	-
Medicaid	6,459	-	-	1,060	-	-	16.41%	-	-

Table 2: Diabetes

	Individuals with diabetes			Prevalence of diabetes based on having at least one outpatient claim		
	VHCURES	Rutland EHR System	<i>Difference</i>	VHCURES	Rutland EHR System	<i>Difference</i>
TOTAL	1,916	1,500	416	8.97%	7.41%	1.56%
Age Group						
25-34	120	78	42	2.67%	1.56%	1.11%
35-44	305	209	96	6.69%	4.98%	1.71%
45-54	585	476	109	9.32%	8.50%	0.82%
55-65	906	737	169	15.02%	13.54%	1.48%
Gender						
Female	949	707	242	7.86%	6.76%	1.10%
Male	967	793	174	10.42%	8.10%	2.32%
Payer						
Commercial	1,349	-	-	9.06%	-	-
Medicaid	567	-	-	8.78%	-	-

Table 3: HbA1c Testing and Diabetes

	Individuals <i>without</i> diabetes among individuals with at least one HbA1c test			Prevalence of <i>no</i> diabetes among individuals with at least one HbA1c test			Individuals <i>with</i> diabetes among individuals with at least one HbA1c test			Prevalence of diabetes among individuals with at least one HbA1c test		
	VHCURES	Rutland EHR System	<i>Difference</i>	VHCURES	Rutland EHR System	<i>Difference</i>	VHCURES	Rutland EHR System	<i>Difference</i>	VHCURES	Rutland EHR System	<i>Difference</i>
TOTAL	972	1,011	-39	38.31%	51.95%	-13.64%	1,565	935	630	61.69%	48.05%	13.64%
Age Group												
25-34	115	163	-48	56.65%	84.46%	-27.81%	88	30	58	43.35%	15.54%	27.81%
35-44	173	208	-35	41.99%	63.80%	-21.81%	239	118	121	58.01%	36.20%	21.81%
45-54	317	307	10	39.77%	49.52%	-9.75%	480	313	167	60.23%	50.48%	9.75%
55-65	367	333	34	32.62%	41.26%	-8.64%	758	474	284	67.38%	58.74%	8.64%
Gender												
Female	530	546	-16	40.83%	55.26%	-14.43%	768	442	326	59.17%	44.74%	14.43%
Male	442	465	-23	35.67%	48.54%	-12.87%	797	493	304	64.33%	51.46%	12.87%
Payer												
Commercial	731	-	-	39.32%	-	-	1,128	-	-	60.68%	-	-
Medicaid	241	-	-	35.55%	-	-	437	-	-	64.45%	-	-

Appendix A – Zip Codes in Rutland Catchment Area

The Rutland catchment area is defined as zip codes serving at least 400 Community Health Centers of the Rutland Region (CHCRR) patients, with two exceptions noted below.

05701
05733
05735
05736* – Rutland Center
05738
05739
05743
05744* – Florence
05757
05759
05761
05763
05764
05765
05773
05774
05777

(*less than 400 CHCRR patients but entirely surrounded by other zip codes with 400 or more CHCRR patients)

Vermont Million Hearts Report

Examining the Utility of Vermont's All Payer Claims Data base

Vermont Million Hearts: Case Study #2 Report

Vermont Million Hearts: Case Study 2 Report

Objective

The original purpose of this case study was to generate and compare population health care statistics obtained from two data sources: the Vermont Healthcare Claims Uniform Reporting Evaluation System (VHCURES) and the Vermont Uniform Hospital Discharge Data Set (VUHDDS). We used data from each source to compare the incidence of heart attack, stroke, and fatal stroke in the Vermont population between 2008 and 2012, broken out by age, gender, and county. For heart attack and stroke, we also estimated expenditures per event.

We highlight key results from this analysis and discuss reasons for the differences in results generated by the two data sources. We had hoped that quantifying the differences in findings by data source would provide insight into how the two data sources (VHCURES and VUHDDS) can give Vermont policymakers a fuller picture of health status than can either of the two individually. Instead our analysis provided lessons on the challenges associated with this undertaking and some recommended next steps for future research.

Data Sources

VHCURES, Vermont's all-payer claims database, includes health insurance enrollment and claims data for Vermont residents covered by insurance companies with more than 200 enrollees in Vermont. It excludes those covered by TRICARE, the Federal Employees Health Benefit Plan (FEHBP), or the Veterans Administration (VA); Medicare enrollees not also covered by Medicaid; and the uninsured. For this case study, we used the December 2013 release of VHCURES released by Onpoint Data Systems, Portland, Maine, which is limited to enrollment and claims data from Medicaid and commercial carriers.

By contrast, VUHDDS has a mandate to include discharge data from Vermont general acute care hospitals, as well as data for all Vermont residents discharged from hospitals in New York, New Hampshire, and Massachusetts. Discharge data from mental health hospitals and nursing facilities are not included in VUHDDS.

There are some notable differences in these data sources that can affect incidence rates.

- **Insurance status.** VUHDDS includes data on individuals not captured by the VHCURES data—those covered by TRICARE, FEHBT, or the VA; Medicare enrollees not also covered by Medicaid; and the uninsured.
- **Settings of care.** VHCURES contains information about care provided in a broader range of settings (e.g., freestanding clinic, MD office, acute care inpatient, home health, nursing, hospice, transportation). VUHDDS includes only information about care

provided at settings associated with acute-care facilities (e.g., hospital-based inpatient, hospital-based ambulatory care).

- **Facilities.** VHCURES includes information about all hospitals visited by Vermont residents, whereas VUHDDS is limited to information about hospitals in Vermont and bordering states.
- **Reporter.** VUHDDS contains the discharge data as reported by the facility, while VHCURES claims come from the payer (or payers) who receive the insurance claims for reimbursement purposes.
- **Unique person identifier.** VHCURES allows one to track the care given to a unique person across time and/or facility. VUHDDS data is on an encounter level with no unique person identifier.

Both VHCURES and VUHDDS contain discharge data that include diagnosis codes, patient age, gender, procedure codes, revenue codes, and total charges associated with a visit.

Inclusion Criteria

We defined the population included in our analyses by age, gender, county, and insurance plan. We used the following criteria to minimize the distortions that could be introduced by the differences between the data sources noted above.

- **Age criterion.** To eliminate issues associated with newborn complications and distortions due to lack of Medicare enrollees in VHCURES, we limited the population to Vermont residents between the ages of 1 and 64 for each year of analysis.
- **Residence criterion.** We limited the VHCURES and VUHDDS population to Vermont residents.
- **Facility criterion.** We ran the VHCURES data two ways: with all event claims and with inpatient and emergency department claims only, so that the data would be comparable to those run with VUHDDS data. We predicted that limiting the VHCURES sample to emergency department and inpatient events would eliminate potential double counting from physician billing and other services for the same event and align with VUHDDS, which has only data from hospital stays.
- **Behavioral health exclusion.** We excluded individuals in VHCURES enrolled in behavioral health only plans to avoid double counting in the denominator, because most of these individuals also were listed as having medical care coverage. All heart attack and stroke events were included in the claims count, regardless of payer.

Heart attack and stroke rates and fatal strokes were calculated by county. We used U.S. Census data of insured individuals as the denominator for VUHDDS and actual enrollment to estimate the denominator for VHCURES.

Results

Despite the adjustments that we made to minimize distortions, we were not able to generate comparable events from the VHCURES and VUHDDS databases. The VHCURES “all event” counts were roughly double those from VUHDDS, whereas the VHCURES “inpatient and emergency department only” events were only half of those reported in VUHDDS. We believe that these differences occurred for three reasons:

1. The VUHDDS data are made up of hospital discharge data only and do not include other charges such as for ambulance and long-term residential care
2. The VHCURES variables that we chose to select facilities and emergency room visits did not adequately capture all acute care inpatient claims using our current specifications and were possibly too highly reliant on payer-specific coding practices
3. The VHCURES data were not limited to acute care inpatient facilities in Vermont

We also compared the VUHDDS data with the Healthcare Cost and Utilization Project (HCUP) data and found the resulting numbers from these two discharge databases to be comparable, which suggests that the numbers reported in VUHDDS are accurate and that further research is needed to replicate these numbers in VHCURES.

Regardless of the differences in numbers, both databases did show that heart disease and stroke are significant problems for Vermont residents, regardless of definition. We discuss the utilization as reported by VHCURES and VUHDDS and the expenditures as reported in VHCURES. VHCURES expenditure amounts are tabulated separately from both the “all event” and “inpatient and emergency department only” facility restriction variations. These alternative expenditure calculations may be viewed as representing the range in payments for heart disease, because not all care is confined to the inpatient setting.

Heart Attack

As mentioned above, the counts for heart attack vary significantly, depending on the data source. However, regardless of the source, the data show that heart attack is an important public health issue in Vermont (see Table 1). As expected, incidence of heart attack increased markedly with age, as did the “allowed amount.” Allowed amount is our proxy for price and is calculated by adding the amount paid by the insurer and the out-of-pocket expenses such as copayment and deductible paid by the enrollee. The allowed amount increased steadily between 2008 and 2010, but in both of the VHCURES samples, it increased sharply between 2010 and 2011 and then dropped between 2011 and 2012. There was roughly a 3:1 ratio of heart attacks among males compared with females, for both VHCURES samples and VUHDDS. Allowed amount for commercial payers was roughly twice that for Medicaid in both VHCURES samples.

Stroke

Table 2 shows that the results from the stroke data are quite different from those for heart attack. Although the increase in incidence of stroke with age was similar to that for heart attack, except for the significant increase in incidence of stroke in the 55-64 age group in the VHCURES all event data, the incidence of stroke was similar among males and females, whereas heart attack rates were much higher among males. Of particular note is the fact that the VHCURES emergency department and inpatient data yielded allowed amount estimates that were between two and six times those calculated from the VHCURES all events data, due in all likelihood to the relative expense of an inpatient stay compared with that of a follow-up visit with a physician or rehabilitation. Interestingly, the allowed amount for the 18–34 age group from the VHCURES emergency room and inpatient sample was very high and well above the overall average for that sample. As with heart attack, allowed amount for stroke events was higher for commercial payers than for Medicaid.

Fatal Stroke

The patterns in the lack of comparability in the three samples carried over to the fatal stroke analysis. Despite the significant increase in incidents of stroke in the 55–64 age group, there was not a related significant increase in fatal strokes for that age group. The incidence of fatal stroke was comparable across genders in the VUHDDS sample, but for the VHCURES samples, there was a significantly higher incidence among men.

Conclusions and Next Steps

Both VHCURES and VUHDDS confirmed important facts: heart disease is a common reason for hospital inpatient admissions, a condition which increases with age, and the utilization associated with this condition makes it a driver of health care expenditures.

However, though these overall findings were similar between the two data sources, the incidence reported by the two sources was very different, was difficult to reconcile and demonstrated the challenges in comparing health statistics across sources. This lack of alignment sheds light on the strengths and weaknesses of each source for further research, both for this study and for other analyses which use these databases.

For this particular study, some of the problems we encountered can be addressed by refinement of the specifications. For example, for this study, we recommend that the VHCURES data be limited to 14 acute-care inpatient facilities in Vermont, using the master provider index in the VHCURES data. We further recommend that for similar analyses, inpatient visits be selected using final-status claims as identified by claims with a valid room and board charge at acute-care facilities using the method that Truven Health MarketScan® deploys when working with multipayer data.

However, even with these restrictions, we caution that the numbers from the two data sources could still not exactly match. The VUHDDS data may not have the correct primary payer listed

on the discharge form if the primary payer changed since the claim was filed (in other words, the primary payer listed on a discharge form may change after the original claim filing). This would likely impact individuals covered by Medicare and another carrier (e.g., Medicaid or commercial) which may cause inadvertent inclusion or exclusion. Additionally, VHCURES will not include the experience for the uninsured, self-pays, or people covered by plans (such as Medicare) which are not required to report to VHCURES.

In general, we would advise health researchers working with VUHDDS and VHCURES to remember the strengths and weaknesses of the underlying databases when conducting analyses. VUHDDS is encounter-based and covers all acute-care facilities in Vermont and is ideal for showing inpatient care and emergency room utilization trends at a facility, regardless of patient residence, insurance status or payer. VHCURES, by contrast, is payer-based and covers nearly all settings of care and is useful when looking at overall trends of healthcare system utilization by insured Vermonters.

Tables 1–3 that follow contain data on the incidence of and expenditures for heart attack, stroke, and fatal stroke among Vermont residents from VUHDDS and the two VHCURES samples discussed above. Despite the differences in the incidence between the two sources, examining the tables does give a sense of the importance of heart disease as a public health issue. Further study using these two sources could shed additional light on the utilization and expenditure patterns as well as further reconciling the findings from the two data sources.

Table 1. Heart Attack Incidence and Expenditures Among Vermont Residents—VHCURES All Events, VHCURES ED and IP Only, and VUHDDS, 2008-2012, by Age, Gender, Payer, and County

		Total Count			Total Rate		Total Allowed Amount	
		VHCURES All Events	VHCURES ED and IP Only	VUHDDS	VHCURES ED and IP Only	VUHDDS	VHCURES, All Events	VHCURES, ED and IP Only
TOTAL		4,447	856	2,019			\$10,226	\$13,471
Year								
	2008	796	192	393	4.0	7.7	\$9,912	\$11,435
	2009	820	179	364	3.6	7.2	\$9,438	\$11,927
	2010	1,064	152	464	3.1	9.1	\$9,754	\$13,184
	2011	850	150	372	3.1	7.3	\$11,147	\$18,748
	2012	917	183	426	3.8	8.4	\$10,898	\$13,029
Age Group								
	1-17	*	*	*	*	*	*	*
	18-34	*	*	*	*	*	*	*
	35-44	415	125	205	3.5	5.6	\$8,458	\$8,303
	45-54	1,395	273	669	6.0	14.1	\$10,448	\$16,254
	55-64	2,573	445	1,110	10.5	24.7	\$10,473	\$13,371
Gender								
	Female	1,155	188	539	1.5	4.2	\$9,108	\$12,074
	Male	3,292	668	1,480	5.6	11.9	\$10,618	\$13,864
Payer								
	Commercial	3,327	644	1,536	3.7	8.3	\$11,817	\$15,564
	Medicaid	1,120	212	483	2.5	6.9	\$5,500	\$7,112
VT Counties								
	Addison	185	41	163	3.1	10.9	\$12,701	\$14,011
	Bennington	411	61	73	4.4	5.1	\$8,740	\$13,137
	Caledonia	188	33	65	2.8	5.3	\$10,335	\$10,935
	Chittenden	548	154	369	2.4	5.5	\$14,868	\$18,090
	Essex	90	*	*	*	*	\$7,059	\$13,369
	Franklin	364	77	245	4.3	12.2	\$7,703	\$11,096
	Grand Isle	37	*	*	*	*	\$10,932	\$7,380
	Lamoille	148	14	106	1.4	10.9	\$14,908	\$53,849
	Orange	235	37	74	3.5	6.3	\$8,701	\$11,364
	Orleans	366	57	121	5.7	11.9	\$7,486	\$6,291
	Rutland	622	136	273	5.6	11.1	\$10,017	\$11,887
	Washington	335	65	282	2.6	11.5	\$11,823	\$14,450
	Windham	470	69	98	4.0	5.7	\$8,746	\$10,433
	Windsor	448	92	104	4.2	4.7	\$9,611	\$12,108

ED, emergency department; IP, inpatient; Total Rate, events per 10,000 enrollees with medical insurance.

* Suppressed due to cell size.

Table 2. Stroke Incidence and Expenditures Among Vermont Residents—VHCURES All Events, VHCURES ED and IP Only, and VUHDDS, 2008-2012, by Age, Gender, Payer, and County

		Total Count			Total Rate		Total Allowed Amount	
		VHCURES All Events	VHCURES ED and IP Only	VUHDDS	VHCURES ED and IP Only	VUHDDS	VHCURES All Events	VHCURES ED and IP Only
TOTAL		9,091	288	1,047			\$3,313	\$11,685
Year								
	2008	1,511	50	195	1.0	3.8	\$3,624	\$8,596
	2009	1,658	60	186	1.2	3.7	\$3,544	\$11,114
	2010	1,910	74	237	1.5	4.7	\$2,958	\$10,113
	2011	2,049	50	216	1.0	4.2	\$3,196	\$12,056
	2012	1,963	54	213	1.1	4.2	\$3,346	\$16,989
Age Group								
	1-17	145	*	*	*	*	\$6,207	\$1,423
	18-34	807	*	*	*	*	\$2,245	\$13,723
	35-44	749	44	93	1.2	2.6	\$3,202	\$8,932
	45-54	1,889	75	328	1.7	6.9	\$4,208	\$15,103
	55-64	5,501	150	561	3.5	12.5	\$3,101	\$10,606
Gender								
	Female	4,474	79	428	0.6	3.3	\$3,153	\$13,783
	Male	4,617	209	619	1.8	5	\$3,468	\$10,892
Payer								
	Commercial	3,633	213	767	1.2	4.2	\$6,044	\$11,961
	Medicaid	5,458	75	280	0.9	4	\$1,495	\$10,901
VT Counties								
	Addison	549	13	84	1.0	5.6	\$3,094	\$14,500
	Bennington	378	13	56	0.9	3.9	\$4,211	\$9,365
	Caledonia	669	15	31	1.3	2.5	\$1,600	\$10,804
	Chittenden	1,900	43	267	0.7	4.0	\$3,482	\$20,550
	Essex	137	*	*	*	*	\$1,260	\$5,411
	Franklin	705	24	125	1.3	6.2	\$3,277	\$9,204
	Grand Isle	110	*	*	*	*	\$2,933	\$10,125
	Lamoille	353	12	52	1.2	5.3	\$4,223	\$7,141
	Orange	462	*	22	*	1.9	\$3,298	\$13,937
	Orleans	891	*	45	*	4.4	\$1,155	\$2,208
	Rutland	960	34	128	1.4	5.2	\$3,471	\$11,891
	Washington	509	29	104	1.2	4.2	\$5,544	\$10,754
	Windham	678	25	64	1.4	3.7	\$3,942	\$8,247
	Windsor	790	65	52	2.9	2.4	\$4,389	\$9,840

ED, emergency department; IP, inpatient; Total Rate, events per 10,000 enrollees with medical insurance.

* Suppressed due to cell size.

Table 3. Incidence of Fatal Stroke Among Vermont Residents—VHCURES All Events, VHCURES ED and IP Only, and VUHDDS, 2008-2012, by Age, Gender, and Payer

		Total Count			Total Rate*	
		VHCURES, All Events	VHCURES ED and IP Only	VUHDDS	VHCURES ED and IP Only	VUHDDS
TOTAL		66	16	38		
Year						
	2008	19	*	*	*	*
	2009	14	*	*	*	*
	2010	11	*	10	*	0.2
	2011	12	*	*	*	*
	2012	10	*	*	*	*
Age Group						
	1-17	*	*		*	*
	18-34	*	*	*	*	*
	35-44	*	*	*	*	*
	45-54	19	*	13	*	*
	55-64	41	10	23	0.2	0.5
Gender						
	Female	26	*	19	*	0.1
	Male	40	*	19	*	0.2
Payer						
	Commercial	38	*	23	*	0.1
	Medicaid	28	*	15	*	0.2

ED, emergency department; IP, inpatient; Total Rate, events per 10,000 enrollees with medical insurance.

* Suppressed due to cell size.

Vermont Million Hearts Report

Examining the Utility of Vermont's All Payer Claims Data base

Linking Data sources in Vermont: Pros, Cons, and Alternatives

Linking Data Sources in Vermont: Pros, Cons, and Alternatives

Introduction

The purpose of this memo is to describe how data sources in Vermont can be used in combination to achieve a better understanding of the underlying population. We outline the types of information that can be gleaned by combining data from different sources, the benefits and drawbacks of data combination, and alternatives to data combination that can still inform health researchers.

Background

Researchers use a variety of data sources to study population health in Vermont. Those sources include surveys, registries, and claims and discharge data.

- Surveys contain self-reported responses to questions; survey data are often weighted to reflect the Vermont population. Surveys are often conducted at annual or biannual intervals.
- Registries contain information intended to show a real-time snapshot of the population. Information for registries is collected frequently.
- Claims and discharge data are based on billing information for visits to a health care provider, but they contain different types of information. Claims data contain information on what insurers paid for a particular service, regardless of location or type of care. Discharge data contain information about a facility-based inpatient stay and, for Vermont, some emergency department (ED)¹ visits and observational stays. Note that claims and discharge data may not provide a complete picture of what occurs during a visit to a health care provider but rather may reflect administrative data used for insurance billing purposes.
- Electronic health record (EHR) data is used increasingly by practices throughout the nation to reflect the details contained in a patient's medical record. These data are detailed and are considered the gold standard of care provided to a patient during a health care encounter.
- Supplementary data on the underlying health care characteristics by county are collected by the U.S. government. These data can contain information such as population estimates by age cohort, education level, number of health care providers in the area, and unemployment rate.

The following are examples of some of these data sources:

¹ A list of acronyms is provided in Appendix A.

- The Vermont Healthcare Claims Uniform Reporting and Evaluation System (VHCURES) contains health insurance claims data that include de-identified eligibility records and medical and pharmacy paid claims for more than 90 percent of the Vermont population.
- The Vermont Uniform Hospital Discharge Data Set (VUHDDS) contains data on all inpatient discharges submitted by Vermont hospitals and includes diagnosis codes and procedure codes. These data also include information on Vermont residents whose discharges occurred in border states such as New Hampshire or Massachusetts.
- The Behavioral Risk Factor Surveillance System (BRFSS) contains data generated by a telephone survey completed annually by a representative sample of Vermont residents.
- The Pregnancy Risk Assessment Monitoring System (PRAMS) contains data collected by paper survey with telephone follow-up of Vermont mothers who recently had a live birth. The data collected include information on preconception health and family planning and prenatal care.
- The Vermont Cancer Registry (VCR) is updated regularly as cancer diagnoses occur. It contains information such as demographic factors and primary payer on all Vermont residents who have been diagnosed with cancer.
- The Vermont Immunization Registry (IMR) contains data on the date and type of vaccine given to all individuals born in Vermont who had a vaccine in a provider practice or who were vaccinated out of state and made an insurance claim through a Vermont provider. It is updated on an ongoing basis.
- Births, deaths, and other vital statistics include incidents that occurred in Vermont and those that occurred to Vermont residents in other states.
- National registries include sources such as the Women, Infants, and Children (WIC) nutrition program data.
- The Area Health Resource File is a publically available source of yearly data outlining health care market characteristics and sociodemographics by county.

Appendix B contains a detailed listing and description of the data sources that are available to public health researchers in Vermont.

Combining Data Across Sources

To understand multiple dimensions of the population, public health researchers are increasingly interested in combining data from separate sources. This practice, we refer to in this document as *linking*, can allow researchers to conduct population-based analyses that would otherwise not be possible. Such analyses can yield an enhanced health care picture that is more detailed than that obtainable from a single source.

An example of valuable data linkage is the Massachusetts Pregnancy to Early Life Longitudinal (PELL) Data System, which links hospital discharge data with birth, stillbirth, and death certificates. This linkage has allowed Massachusetts to conduct population-based analyses in a number of areas—including the impact of adequate prenatal care, the effect of gestational diabetes during subsequent pregnancies, factors involved in cesarean deliveries, portraits of care provided prior to maternal death, and the impact of early intervention programs.

Although combining data is sometimes useful in providing a fuller picture of care, it should be approached with some caution. Combining data from several sources, even if these sources alone are de-identified, can cause the *mosaic effect*, whereby combining data can turn previously anonymous data into personally identifiable data and compromise patient confidentiality. Given the sensitivity of health data and under the guidance of Health Insurance Portability and Accountability Act (HIPAA) regulations, it is essential to limit exposure to such risks.

Additionally, there are technical considerations associated with linking data. First, some data sources simply cannot be linked because they may not have any common fields. For example, it would make no sense to try to link survey data based on a sample of the population with other data that are based on the entire population. Second, linking some data sources can introduce bias into the resulting linked database. For example, in the all-payer claims data, only people with coverage are present, leaving out the experience of those lacking health insurance. This problem is compounded when these claims are combined with the WIC data, in which women with more stable home lives are likely to be present in any linked database because of matching information between the two sources and the experience of women and children with less stable home lives are not as well captured because of incomplete or conflicting data between the two data sources. Third, linkages can be very time consuming to program and review for accuracy. This is not a “push button” approach—one needs to analyze whether the correct records are being matched to one another and to access and quantify any bias in the linked records compared with unlinked records to avoid drawing incorrect conclusions from the linked database.

Methods of Combining Data Across Sources

Two techniques are used for linking different data sources: *deterministic* matching and *probabilistic* matching. Deterministic matching takes two data sources and uses common elements between the two data sources to join them. An example of deterministic matching would be connecting the VHCURES data with the Area Health Resource File (AHRF) to find the demographics for a given county (e.g., total population, total population older than 65, number of households) using the field “county,” which is present in both datasets. Such a combination would allow one to look at trends in health care expenditure by education level, for example, to assess whether public health messages are appropriately written for all population groups.

Probabilistic or “fuzzy” matching is performed only after deterministic matching has been completed and identified matches have been removed from the data sources. Probabilistic matching attempts to match those records that were left unmatched after the initial deterministic (one-to-one) matching process have been identified using incomplete information. For example, if one were matching births as recorded in vital statistics records to discharge data with identifiable personal-level information, one would first deterministically match using all fields that these two data sources have in common: medical record number, birth certificate number, date of birth, age of mother, sex of baby, and facility. In cases where some of these fields are missing, as is often the case with administrative data, one would match the remaining records using a subset of these fields such as birth certificate number, date of birth, and sex of baby.

These attempts to match the records using imperfect information typically drop the most restrictive criterion first, and then drop the second most restrictive criterion, and so on. Records matched during each iteration are removed from the pool of unmatched records for subsequent matching. These matches must be carefully assessed to determine whether the correct records from one database are combined with the correct records from the other database. Sometimes, if highly identifiable information subject to different spelling is used for matching (e.g., names), one can use algorithms that match on the sound of the name or on common abbreviations.

Combining data from two sources can yield important information that is unavailable from a single data source alone. However, probabilistic matching and review should be done only by an experienced programmer, data analyst and subject expert working in conjunction. The experience team has to ensure that the matches are legitimate for the subject being examined and that any bias in the population found in the linked database that was not in the unlinked records is quantified before proceeding with any analysis.

Assessing the Appropriateness of Linkage

In this section, we assess which data sources used by the Vermont Department of Public Health are amenable to linkage (see Appendix A for a list of those data sources). Linking across data sources should be approached with caution. Linking some types of data sources provides clear benefits, linking others may be beneficial but with caveats, and linking some others is ill-advised. Some data sources may not be amenable to linking, and linking others may bias the data in ways that cannot be anticipated.

One major drawback of probabilistic matching is that not all records can be matched, even with highly identifiable data. Individuals make mistakes in coding fields, names change, and people move. Therefore, it is critically important to compare the linked database with the unlinked records to assess any underlying bias in the records that cannot be linked using either probabilistic or deterministic techniques. In our experience, such bias often represents information about the care given to the underserved that may not be correct because of transient living situations, imperfect communication due to language, or inability to communicate due to underlying medical issues.

First, we discuss data sources that are well suited for linkage. These data sources contain files that have standard definitions and high levels of accuracy and contain fields that are either highly identifiable or accurate (e.g., birth certificate number, which is a legal record) or fields that are standardized (e.g., county code, which has a common definition regardless of the data source).

Below are some examples of data sources that may make sense to combine:

- Linking vaccination program registry data with birth certificate data can yield information on the success of the program for babies and children and highlight any domains where targeted efforts could be directed (e.g., certain geographic areas, sociodemographic groups).

- Linking any geographically based census data with data from the All Payer Claims Database (APCD) can allow one to look at trends while controlling for the underlying sociodemographics of the area in which the patient lives.
- Linking general facility information data such as bed size with claims data could provide insight into hospital utilization by capacity. For example, one could find that a high number of emergency department (ED) visits may be associated with an underlying lack of substance abuse beds in local facilities.

Some data sources may be well suited for combination but with some caveats because of bias in the records that match across data sources compared with those records that do not match:

- Linking hospital discharge data with birth certificates allows for population-based analyses of infants and mothers. Both are based on birth and delivery events as well as defining characteristics such as date of birth, location of birth, location of mother, and sex of child. The drawbacks involve incorrect linkages for births that share all demographic characteristics in the claims data and underrepresentation of the underserved population in the linked file. As always, careful review is needed before proceeding with analyses.
- Linking VHCURES claims data and hospital discharge data allows for analyses of outpatient care following discharge from the hospital. Matching data between these two sources is promising, because both are event-based administrative claims that can be matched using payer, diagnosis, facility, length of stay, age, sex, and Zip code. However, both data sources have limitations: the discharge data has no person identifier, and VHCURES does not cover the experience of patients who are insured by payers covering fewer than 200 Vermont residents, Federal health plan employees, or the uninsured. Moreover, VHCURES has an imperfect person identifier for individuals who are covered by separate insurers for medical, behavioral health, and pharmacy coverage. Both of these limitations will affect types of research one could do with the linked database.
- Linking VHCURES data with death certificates allows for analyses of treatment prior to death. In addition to the difficulty of tracking individuals with different insurance coverage mentioned in the previous bullet, combining data accurately can be difficult for those who die outside the hospital setting because of imperfect record-keeping protocols in non-facility-based claims.
- Linking discharge data to Consumer Assessment of Healthcare Providers and Systems (CAHPS) data has the advantage of controlling for the effect of patient satisfaction. However, there is some uncertainty as to whether the facilities and providers are the same in these two data sources, so researchers should carefully examine results when linking these two sources.

Finally, combining some data sources is either ill-advised or recommended only for research with major caveats:

- Attempting to link claims or discharge data with any of the individual health surveys (e.g., PRAMS, Behavioral Risk Factor Surveillance System [BRFSS], Adult Tobacco Survey [ATS], Asthma Call-Back Survey [ACBS]) presents challenges because of the lack of common data elements between data sources. Also, linking any data source with a

survey may not yield a productive database and might be better examined by comparing the results from the two data sources separately.

- Linking claims data to Vermont Household Health Insurance Survey (VHHIS) data has the advantage of capturing the effects of insurance status with claims. However, depending on the accuracy of the location information provided and the fact that only people with a landline are contacted, the results could be biased and would certainly compromise patient confidentiality.

Alternatives to Linking

Given the limitations and hazards of linking data across sources described above, we recommend against linking unless there is a clear research need. Examples of such needs include characterizing the sociodemographic and health care environment of a geographical area to include in multivariate analysis. Instead, we recommend generating tables/figures from complementary data source separately and interpreting findings in consideration of both perspectives. A side by side comparison can also be used to determine whether similar patterns are observed across multiple data sources despite data collection or inclusion differences. This can provide researchers or policymakers with evidence that the statistics or trends observed are generalizable across populations or data collection methods.

When comparing two data sources, one should use best practices, including:

- Having a detailed analysis plan outlining the purpose of the study, including hypotheses and any known shortcomings of each underlying data source
- Providing all information needed to understand the data source
- Defining any outcomes or descriptive data elements in the same way for each data source. For example, comparing health claims data to EHR data should look at the same time period, have similar geographic areas and place of service, and have the same specifications regarding diagnostic codes used for any measures
- Providing tables that compare results by source, in a side-by-side basis. This approach can make a compelling case without the potential bias, cost, and risks associated with linkage.

Summary

Combining data sources that are available to public health researchers is a powerful and increasingly used technique that requires careful thought and detailed review by expert technicians, has the ability to breach patient confidentiality, and often contains some bias. Although certain research questions can be answered only by using such a method, looking at the sources separately is the best practice for gauging underlying health care issues and trends.

APPENDIX A

List of Acronyms

APCD	All Payer Claims Database
AHRF	Area Health Resource File
ATS	Adult Tobacco Survey
ACBS	Asthma Call-Back Survey
BRFSS	Behavioral Risk Factor Surveillance System
CAHPS	Consumer Assessment of Healthcare Providers and Systems
EARS	Early Aberration Reporting System
eHARS	Electronic HIV/AIDS Reporting System
HER	Electronic health record
ED	emergency department
EPHT	Environment Public Health Tracking
HIPAA	Health Insurance Portability and Accountability Act
IMR	Immunization Registry
NEDSS	National Electronic Disease Surveillance System
PELL Data System	Pregnancy to Early Life Longitudinal Data System
PRAMS	Pregnancy Risk Assessment Monitoring System
SATIS	Substance Abuse Treatment Information System
STDMIS	Sexually Transmitted Disease Management Information System
VCR	Vermont Cancer Registry
VHCURES	Vermont Healthcare Claims Uniform Reporting and Evaluation System
VHHIS	Vermont Household Health Insurance Survey
VPMS	Vermont Prescription Monitoring System
VUHDDS	Vermont Uniform Hospital Discharge Data Set
WIC	Women, Infants, and Children
YRBS	Youth Risk Behavior Survey

APPENDIX B

Data Encyclopedia



Health Surveillance

Data Encyclopedia

A review of data sources and resources available at the Vermont Department of Health

Relationships to the Vermont Health Care Innovation Project performance measures

Dayman, Caitlyn
6/19/2013

The Division of Health Surveillance, Public Health Statistics, has recently compiled the "Data Encyclopedia: A Review of Data Sources and Resources Available at The Vermont Department Of Health." This publication provides an overview of the commonly used data sources to assess and track population health outcomes and contributors to disease in Vermont. It is intended to provide a high level description of the type of information in each data source, the potential uses and limitations of the data, and the existing reports summarizing the data. For additional information on accessing data from these sources, generating reports and interpreting the significance, please contact Caitlyn Dayman (Caitlyn.dayman@state.vt.us) at the Vermont Department of Health.

Contents

Data Sources vs. Data Resources	37
Data Sources:.....	39
Surveys.....	39
<i>Adult Tobacco Survey (ATS)</i>	<i>39</i>
<i>Behavioral Risk Factor Surveillance System (BRFSS)</i>	<i>40</i>
<i>Asthma Call Back Survey (ACBS)</i>	<i>41</i>
<i>Consumer Assessment of Healthcare Providers and Systems (CAHPS)</i>	<i>42</i>
<i>Health Care Provider Surveys</i>	<i>43</i>
<i>Pregnancy Risk Assessment Monitoring System (PRAMS)</i>	<i>44</i>
<i>School Nurses' Report.....</i>	<i>45</i>
<i>Youth Risk Behavior Survey (YRBS)</i>	<i>46</i>
Registries and Surveillance Systems.....	47
<i>Covisent Docsite</i>	<i>47</i>
<i>EARS (Early Aberration Reporting System).....</i>	<i>48</i>
<i>Electronic HIV/AIDS Reporting System (eHARS).....</i>	<i>49</i>
<i>Infectious Disease Outbreak Database.....</i>	<i>50</i>
<i>National Electronic Disease Surveillance System (NEDSS) aka NEDSS Base System (NBS)</i>	<i>51</i>
<i>Refugee Health data system</i>	<i>52</i>

<i>SATIS – Substance Abuse Treatment Information System</i>	53
<i>STDMIS - STD surveillance dataset</i>	54
<i>Vermont Cancer Registry</i>	55
<i>Vermont Immunization Registry (IMR)</i>	56
<i>Vermont Lead Database</i>	57
<i>Vermont Prescription Monitoring System (VPMS)</i>	58
<i>Vital Statistics</i>	59
<i>VMAP Access Database and CAREWare</i>	61
<i>Claims and Discharge data</i>	62
<i>Vermont Healthcare Claims Uniform Reporting and Evaluation System (VHCURES)</i>	62
<i>BluePrint VHCURES data set</i>	63
<i>Vermont Uniform Hospital Discharge Data Set (VUHDDS)</i>	64
<i>Data Resources:</i>	65
<i>Healthy Vermonters 2020 toolkit</i> (includes the ‘Performance Dashboard’ and ‘Maps & Trends’).....	65
<i>Vermont Environment Public Health Tracking Program (EPHT)</i>	67

Data Sources vs. Data Resources

Data Sources

The data sources included in this document include surveys, registries, and claims and discharge data. Most of these data sources are cleaned and available for secondary data analysis by analysts granted access to the data. For the most part, these data sources have been “cleaned”, which means variables have been formatted, appropriately categorized, and weighted as necessary. However, some data sources contain data that is more “raw” in character and prior to being used some steps should be taken to clean and standardize the data for analysis. The three types of data sources presented here are:

- **Surveys**-Surveys contain self-reported responses to questions. Some, but not all, of the surveys included here are from a sample that is then weighted back to reflect the Vermont population. Surveys are usually completed at one point in time (annually, bi-annually, etc.).
- **Registries and Surveillance Systems**-This type of information is collected frequently and continuously. They are meant to show a real time snapshot of the population. In some cases, information is constantly open to change, so it is important to pay attention to time periods of when information is accessed.
- **Claims and Discharge data**-Both of these are based on billing information for visits to a health care provider. Claims data is information based on what an insurer paid for a given service. Discharge data tells us information about a visit to a health care provider based on diagnosis and procedure codes listed at discharge. It is important to note that both of these data sources rely on billing information and in some cases may not entirely describe what occurs during a visit to a health care provider.

Data Resources

Individuals can access information about population health status and contributors to health through two primary resources developed by the Vermont Department. These portals include access to various data sources that in combination can help to better understand health trends, opportunities for health improvement and current actions for health protection.

Healthy Vermonters Toolkit is built on the concepts of *Results Based Accountability*™ and includes:

Population Indicators (such as smoking prevalence) are measures for which the Health Department, with state government and community partners, share responsibility for making change. All Healthy Vermonters 2020 indicators are displayed. The Health Department routinely uses three ways to assess population indicators (such as smoking prevalence) at the local level: by county, by Health Department district office area, and by hospital service area (HSA). Maps & Trends links you to interactive InstantAtlas™ pages, with maps, tables and graphs for all Healthy Vermonters 2020 indicators and goals.

Performance Measures (such as the percentage of smokers registered with the Vermont Quit Network), are measures for which Health Department programs are responsible for the performance of interventions that, over time, will improve health, as reflected in the population indicators (such as reduced smoking prevalence).

Vermont Environment Public Health Tracking

Tracking brings together environmental and public health data in one place to assist in researching possible health threats from environmental exposures such as air pollution and drinking water contaminants. Local, state, regional, and national data will be available through the Vermont Tracking Network. Funded by the Centers for Disease Control and Prevention as part of the [National Environmental Public Health Tracking Program](#), Vermont's Tracking program also links to comparable information from [other states](#) and to national data.

Data Sources:

Surveys

Data Source	Adult Tobacco Survey (ATS)					
Sponsors	http://healthvermont.gov/research/index.aspx#ats The Health Department oversees data collection and analysis Data should be requested through the Health Department's Tobacco analyst					
Most current Year available	2012 is the most recent data set available.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Conducted annually from 2000 through 2008 As of 2008 it was conducted in even calendar years. (2008, 2010, 2012)	The Vermont Adult Tobacco Survey (ATS) is a telephone survey conducted over an eight week period during the fall of the calendar year. The sample includes 2,000 respondents each year: 1,000 each of smokers and non-smokers regardless of telephone type. Beginning in 2008, the survey has been conducted biannually in even years and includes cell phone users. The survey takes approximately 20 minutes to complete.	Non-institutionalized Vermont adult (18 years and older) residents. Historically, the ATS includes over-sampling of both smokers and 18-24 year olds. In accordance with the study design, the results were weighted by gender, age, smoking status, household composition, telephone type, and county in order to compensate for differences between the sample and the overall Vermont adult population.	Ideal for evaluating the effectiveness of Vermont Tobacco Control Program efforts to reduce smoking and increase awareness and knowledge of smoking-related issues.	Several states conduct Adult Tobacco Surveys but each is unique as this evaluation tool is not part of a national survey and data should not be directly compared to that from other states.	<ul style="list-style-type: none"> Tobacco use: screening and cessation intervention 	2010 and 2008 Adult Tobacco Survey reports (In depth report of the survey results)
Indicators for analysis	Broad topic areas: <ul style="list-style-type: none"> Information on quit attempts and smoker confidence Cessation methods General awareness of cessation programs Secondhand smoke perceptions and exposures <ul style="list-style-type: none"> Attitudes about smoking Provider interventions (some media campaign information available through 2010) As of 2012, policy questions included on the survey Trend analysis can but done for most of these factors.					

Data Source	Behavioral Risk Factor Surveillance System (BRFSS)					
Sponsors	http://healthvermont.gov/research/brfss/brfss.aspx The Health Department oversees data collection and cleaning Data should be requested through BRFSS coordinator					
Most current Year available	Data available 1990 to 2013. *In 2009, started including both cellphone and landline phone surveys for adults; cannot report combined measures that use both landline and cell phones sources until 2011 (had to be landline only). *In 2011, the method for weighting sample data changed—we cannot aggregate data from multiple years between the pre 2011 and post 2011 data periods.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Survey is conducted on an annual basis	<ul style="list-style-type: none"> • Random digit dial telephone survey • Surveys are completed for a representative sample of the population • Information is then weighted with a raking procedure starting with 2011 data (a post stratification method of weighting was used pre-2011) 	Vermont residents	<p>This is ideal for looking at risk factors and prevalence of chronic conditions at a population level in Vermont.</p> <p>This is a well-established survey that allows us to look at some trends over time.</p> <p>Data can be compared across states</p>	<p>It is not a census; we take a representative sample of surveys and weight them to represent the entire population of Vermont.</p> <p>Information is self-reported</p>	<ul style="list-style-type: none"> • Influenza vaccination in the last year • Pneumococcal vaccination • Colorectal cancer screening • Breast cancer screening • Falls (BRFSS collects information on falls in past 3 months, and injuries due to fall) 	<p>Chronic Disease data pages (published annually)</p> <p>BRFSS annual report</p> <p>Burden Documents (Injury, asthma)</p>
Indicators for analysis	<ul style="list-style-type: none"> • Chronic Conditions • Preventive Measures • Health Insurance/Access • Risk Behaviors (smoking/drinking/diet/exercise) 					

Data Source	<i>Asthma Call Back Survey (ACBS)</i>					
Sponsors	BRFSS Coordinator- Health Department / Asthma analyst http://healthvermont.gov/research/brfss/brfss.aspx The Health Department oversees data collection and cleaning Data should be requested through the Asthma analyst					
Most current Year available	Data available 1990 to 2012 for adults and 2010 for children. (2011 children data has been collected but is not yet available because data is not weighted). *In 2009, started including both cellphone and landline phone surveys for adults. *In 2011, the method for weighting sample data changed—cannot aggregate data from multiple years between the pre 2011 and post 2011 data periods. *The child ACBS is only collected in odd years starting with 2011.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Survey is conducted on an annual basis as a follow-up to individuals reporting asthma on the BRFSS. It is conducted for both adults and children (children it is only collected in odd years starting with 2011).	<ul style="list-style-type: none"> • If respondent, who has asthma, agrees to a follow-up call at the end of the BRFSS they will be called and asked to participate in the ACBS <ul style="list-style-type: none"> ○ Parents that report a child in the household has asthma who agree to follow-up call will be asked to participate in the child ACBS. Child ACBS asks “most knowledgeable” to respond to questions regarding child’s asthma, in some cases this is the child, but usually it is a parent or guardian. 	VT Residents with Asthma	More details about residents with asthma including: Medication use, risk factors, triggers, and preventative methods	We survey a sample of VT residents with asthma who completed BRFSS and agreed to follow-up calls then weight the data to estimate statewide values. Self-reported data. However collected over the long term and it appears we are seeing that people are self-reporting information consistently		Asthma data pages (published annually) Asthma Burden report (data from 2008-2010)
Indicators for analysis	<ul style="list-style-type: none"> • Asthma control and severity • Preventive Measures for environmental triggers • Medication use for control • Service utilization • Data can be linked back to all variables examined in the BRFSS 					

Data Source	<i>Consumer Assessment of Healthcare Providers and Systems (CAHPS)</i>					
Sponsors	UVM's Vermont Child Health Improvement Program (VCHIP) collects, cleans, and stores the data. <i>There are a variety of these surveys that individual providers and individual healthcare facilities can purchase and collect internally; these may or may not be the same versions as what VCHIP is using.</i>					
Most current Year available	VCHIP has only collected one year of data (2012). The survey is in currently in the collection phase for 2013 (July 2013).					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Thus far the survey has been conducted once and is scheduled to occur again in 2013. The frequency that VCHIP will conduct this survey after 2013 has not yet been decided, as many providers and agencies may purchase the survey and complete it internally.	Forty-eight of the 185 BluePrint practices approached by BluePrint staff elected to participate in the CAHPS survey conducted by VCHIP. VCHIP mailed the survey to randomly selected patients (adults and children) from the 48 BluePrint primary care practices in Vermont over a three month period in 2012. Patients are sent a letter and a copy of the survey twice and asked to return the survey mail. The only follow-up for completing the survey is the second mailing of the survey.	Patients in a primary care setting. Randomly selected patients that returned the paper survey. Patients could be either adult or youth patients at a participating clinic. A parent or guardian completed the survey if the randomly selected patient was under the age of 18.	Information on provider messaging and satisfaction with care among patients in a primary care setting.	The sample is to a sampling bias as we only have the information on those who chose to return the survey.	<i>Most of the Patient Experience goals can be measured with some form of CAHPS survey—comparisons should be made with caution and methodologies for survey selection and survey questions can vary.</i>	Relationship between Medical Home Recognition and Patient Experience Responses for the CAHPS® PCMH Survey (completed by VCHIP).
Indicators for analysis	<ul style="list-style-type: none"> • Access to care • Communication between practice/provider and patient • Self-management support • Office staff • Shared Decision making among adult respondents • Comprehensiveness (adults-questions about emotional and mental wellbeing, children-questions about injury prevention, growth, emotional stability, diet) 					

Data Source	Health Care Provider Surveys					
Sponsors	Information should be requested through the Health Department's Provider Survey coordinator http://healthvermont.gov/research/HlthCarePrvSrvys/HealthCareProviderSurveys.aspx					
Most current Year available	Physicians (MD and DO)—1979, 1996-2010 (even years only) Dentists—1999-2011 (odd years only) Physician Assistants—1998-2012 (even years only) Adv. Practice Nurse Practitioners (APRNS)—1998-2002 (even years only)					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
The Vermont Department of Health has been collecting information on health care providers since 1994. Physicians (MDs and DOs), Dentists and Physician Assistants are surveyed every two years at the time of their relicensing. Surveys of APRN's were conducted by the Health Department in 1998, 2000, and 2002. More recent surveys of APRNs have been conducted by UVM.	Forms were included with the relicensing with the exception of the 2012 MD forms. That census is being conducted separate from the relicensing.	Licensed health care providers (including Physicians (MD and DO), dentists, PAs, and APRNs) who are actively serving Vermont patients. The APRN survey was last conducted by the Health Department in 2002. Residents and fellows are not included in the population of this survey. Individuals that provide remote services (mostly radiologists and pathologists) for Vermonters, from another state are included in the count of providers. However, their work time devoted to VT residents cannot be determined.	Most physicians/providers complete, because the forms are included with their relicensing, and the Health Department follows up with non-respondents. For most of the provider surveys, some trend comparisons can be made from survey to survey.	Self-reported by providers, and not further verified.		Summary and statistical reports are produced for every survey.
Indicators for analysis	Most statistics can also be shown by sub geographies (groups of townships) <ul style="list-style-type: none"> Numbers and FTEs by specialty by geographical areas Demographics of Physicians Years in practice Training location (i.e. where did they go to medical school?) Turnover and future retirement plans 					

Data Source	<i>Pregnancy Risk Assessment Monitoring System (PRAMS)</i>					
Sponsors	http://healthvermont.gov/research/PRAMS/prams.aspx Information should be requested through PRAMS Coordinator at the Health Department. Survey data is collected annually by the Health Department.					
Most current Year available	Data available 2001-2011.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Data is collected on an ongoing basis and analytic files are updated annually.	This is a paper survey with phone follow-up that is mailed to a random sample of Vermont mothers 2-6 months after having a live birth in VT or NH. Drawn from birth certificate data, the sampling fraction is approximately 1 out of 5. Women with low birth weight infants (<2500g) are over-sampled. Data is weighted to be representative of the population.	Vermont resident mothers who have recently had a live birth.	A linkage to the birth certificate means PRAMS builds upon existing information. PRAMS covers topics not available elsewhere: prenatal care content & barriers to quitting smoking, drinking amount, breast-feeding duration, intention of pregnancy and sensitive questions on drug use and domestic violence.	Only includes pregnancies resulting in a live birth. Self-reported data can tend to under report certain health outcomes. A certificate of confidentiality may improve the reporting of questions in sensitive areas. Smallest level of geography: state of Vermont.		Examples of reports produced by PRAMS may be found at the PRAMS web site: Healthvermont.gov/research/PRAMS/Prams.aspx
Indicators for analysis	<ul style="list-style-type: none"> • Preconception health and family planning • Prenatal Care • Alcohol, Tobacco and Drug use • Stress and abuse • Breastfeeding • Sleep environment • Dental Health • Postpartum care 					

Data Source	<i>School Nurses' Report</i>					
Sponsors	The Maternal and Child Health division coordinate data collection and storage. School Liaisons in the Health Department District Offices act as local level support for questions related to the survey content and monitor for completion. Data should be requested through School Nurses' Report Coordinator in the MCH division.					
Most current Year available	Data is currently available for the 2007-08 school year through 2012-13 school year.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Information is collected annually by School Nurses in public schools throughout Vermont.	Information is reported by parents/guardians to the school nurse. The data is collected using web-based survey software then it is compiled and aggregated by the survey vendor. A final report is submitted to the Division of Maternal and Child Health (MCH) at the Health Department and shared with the Health Department school liaison. This is a convenience sample; information may be collected differently at each school site.	Children in school whose parents provided information to the school nurse.	Information on access to health care and insurance coverage for all school age kids (K-12). There is also information on a students' asthma status and the presence of an asthma action plan at school. Final report includes filterable data by Health Department District Office, Supervisory Union/School District, school, and grade.	Methods and collection materials vary across schools.	<ul style="list-style-type: none"> Adolescent well care visit 	Asthma Burden Report Healthy Vermonter Goals related to School age health and oral health
Indicators for analysis	<ul style="list-style-type: none"> Well child visits Dental visits Insurance status Presence of an asthma action plan 					

Data Source	Youth Risk Behavior Survey (YRBS)					
Sponsors	http://healthvermont.gov/research/yrbs.aspx Data collection and management is overseen by the Health Department Data should be requested through the Health Department's YRBS coordinator.					
Most current Year available	The most current data set available is from 2013. The survey is completed by students every other year (survey occurs during odd years since 1993).					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
The survey is conducted in odd years at all public middle and high schools around the state.	Paper Survey (approximately 100 questions for HS and 70 questions for MS) administered during the school day.	Middle and high school age students in Vermont Schools.	It occurs in (almost) all high and middle schools around the state. Weighted data is available at both a statewide and sub state level. Data can be used for national comparisons.	It does not reach children who do not attend school or who were out the day the survey was administered.		YRBS Statewide Summary Report YRBS Local Summary Reports (by county and by school district) YRBS data briefs (approximately 6 published each year)
Indicators for analysis	The YRBS measures demographic factors as well as the prevalence of behaviors that contribute to the leading causes of death, disease, and injury among youth. Including information on: <ul style="list-style-type: none"> • Risk Behaviors (personal safety, substance abuse, seatbelt use, etc) • Diet • Exercise • Mental health questions related to bullying, sexuality, violent behavior • Youth assets • Perception of risk • Sexual Behaviors 					

Registries and Surveillance Systems

Data Source	Covisent Docsite					
Sponsors	Covisent is the sponsor of this clinical registry. Data should be requested from the BluePrint program at the Vermont Department of Health Access.					
Most current Year available	Data collection began at varying times for different sites. More information about this data source will be determined as data is shared with programs at the health department.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
This data is collected as a registry. Some information is transmitted in real time to the registry platform, other data is hand entered on a daily basis.	This is a clinical registry. Data from various sources is compiled into one database. Data includes: clinical visits information, labs, visits with the community health team, tobacco cessation program, and SASH.	Individuals in receiving care from participating providers and/or utilizing various community resources (BluePrint's Healthy Living workshop, working with BluePrint's community health team, SASH, and Smoking cessation).	TBD when full access to an extract is available.	TBD when full access to an extract is available.		<i>TBD</i>
Indicators for analysis	<i>TBD</i>					

Data Source	<i>EARS (Early Aberration Reporting System)</i>					
Sponsors	Data is maintained and tracked by an epidemiologist in the Infectious Disease Section. (As of Jan 2014-Bradley Tompkins)					
Most current year available	2003-2014					
Frequency	Design	Population (units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Data is updated daily, 7 days a week	Designed to capture and analyze recent Emergency department visit data for trends and signals of abnormal activity that may indicate the occurrence of events significant to public health (eg. outbreaks, unusual illnesses)	All Individual ED visits from participating Vermont hospitals (FAHC, CVH, Copley, BMH, NCH, SVMC)	Provides very fast data (within 24 hours) on ED visit activity at half of Vermont hospitals. Covers roughly 75-80% of ED beds in state. No missing data, there is 100% reporting from all participating hospitals. For some hospitals, data goes back to 2003.	EARS system is old and unsupported by its original sponsor, CDC. Is not capable of handling new generation syndromic messaging formats (HL7) that the healthcare industry is widely adopting. Not all Vermont hospitals participate.		Flu surveillance data on VDH webpage
Indicators for analysis	<ul style="list-style-type: none"> ER Visit Date and hospital name Patient age, gender, town and state Chief complaint, diagnosis, disposition 					

Data Source	<i>Electronic HIV/AIDS Reporting System (eHARS)</i>					
Sponsors	The data system is maintained by the epidemiologist in the HIV/AIDS/STD/Hepatitis unit of the Infectious Disease Section.					
Most current year available	Complete data available 1983 through 2013. The system is updated as data is received.					
Frequency	Design	Population (units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
The system is updated as labs and case report forms are received at the health department, several times a week. Data is uploaded to CDC at the end of every month.	HIV is a reportable disease, as is AIDS. HIV viral load measurements (including non-detectable results) are reported as are all CD4 counts that are under 200 cells/mm ³ . eHARS archives case report forms and lab results.	Any person who is a resident of Vermont, was diagnosed in Vermont, or is receiving care in Vermont for HIV or AIDS. We also receive data about VT residents from other states.	All HIV/AIDS diagnoses among people who were either initially diagnosed in Vermont or are receiving their medical care in Vermont as well as health information regarding all viral loads and CD4 counts under 200.	Lag in reporting when lab results are received without case report forms and therefore cannot be added to the system.		Integrated Epidemiologic Profiles for HIV/AIDS Prevention and Care Planning, Summary Reports, grant proposals
Indicators for analysis	<ul style="list-style-type: none"> • Demographic factors (age, sex, race/ethnicity, residence, risk factors for HIV infection) • Diagnostic and treatment info (earliest date, residence, provider, facility, continuation of care) 					

Data Source	<i>Infectious Disease Outbreak Database</i>					
Sponsors	Data is maintained and tracked by an epidemiologist in the Infectious Disease Section. (As of Jan 2014-Bradley Tompkins)					
Most current year available	1999 -2013					
Frequency	Design	Population (units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Database is updated on a monthly basis	Designed to capture descriptive information of infectious disease outbreaks that have occurred in Vermont or that involve Vermonters.	Aggregate numbers of VT residents who are ill as part of outbreaks are reported	Provides quick access to historic outbreak data. Allows quick extraction of data elements for grant reporting	Relies on archaic software. Captures basic information on each outbreak, sometimes lacks in-depth information that is part of more complicated outbreak investigations.		ELC and emergency preparedness grant reports
Indicators for analysis	<ul style="list-style-type: none"> • Number exposed, ill, sent to doctor/ER, hospitalized, dead • Location of outbreak, setting of outbreak • Causative agent • Mode of transmission • Date outbreak started • Lead investigator 					

Data Source	<i>National Electronic Disease Surveillance System (NEDSS) aka NEDSS Base System (NBS)</i>					
Sponsors	Data is maintained and tracked by an epidemiologist in the Infectious Disease Section. (As of Jan 2014-Bradley Tompkins and Chelsea Dubie)					
Most current year available	Jan 2004 through the present					
Frequency	Design	Population (units)	Strengths	Limitations	VCHIP Measures	Reports citing data source
The system is updated daily as disease reports are received from health care providers and laboratories.	The data is organized by unique occurrences of a reportable disease. Thus individuals could be in the database multiple times due to a diagnosis of different reportable diseases or due to a re-occurrence of the same reportable disease. This system is used by PHNs in District Offices and transmits data electronically to CDC.	Every instance of a reportable disease diagnosed in Vermont. Occurrences in Vermont are represented in this data set (VT residents and out-of-state residents diagnosed in VT). We also receive data on VT residents diagnosed with reportable diseases in other states.	It is a complete surveillance database of all reportable diseases. Allows for analysis of trends over time Analysis can be performed by individual or by disease occurrence.	Some VT residents who are diagnosed out of state may not be reported to VDH.		CDC's Morbidity and Mortality Weekly Report (MMWR) summarizes national reportable disease data, including VT data.
Indicators for analysis	<ul style="list-style-type: none">• Demographic factors (age, sex, race)• Disease-specific data					

Data Source	<i>Refugee Health data system</i>					
Sponsors	Data is maintained and tracked by an epidemiologist in the Infectious Disease Section. (As of Jan 2014-Susan Schoenfeld, Debra Kaigle)					
Most current year available	October 2012 through the present					
Frequency	Design	Population (units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Data system is updated as Domestic Health Assessment forms are received from health care providers doing these initial medical exams. Data are reported to the Office of Refugee Resettlement (ORR) every 4 months.	Data elements in VT's report are based on this standard. ORR identifies best practice elements of an initial medical exam for newly arriving refugees, and requires states to ensure the exams occur. These visits typically occur within 90 days of refugees' arrival in the U.S.	All individuals resettled in Vermont as refugees.	Only source of data that provides a state-based assessment of the health of newly arrived refugees.	1. Lag time between exams and receipt of some reports may be considerable. 2. As a relatively new data source, report capabilities are still being developed.		Trimester reports to the Office of Refugee Resettlement
Indicators for analysis	<ul style="list-style-type: none">• Demographic factors (age, sex, country of origin)• Summaries of diseases of public health importance (e.g., tuberculosis infection, lead levels, hepatitis B status)					

Data Source	SATIS – Substance Abuse Treatment Information System <i>Description of the data collected is here: http://healthvermont.gov/adap/grantees/documents/SATIS_ProviderDataElements.pdf</i>					
Sponsors	ADAP – Anne Van Donsel					
Most current Year available	Data available – approximately FY2000 through FY2013					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Providers submit data monthly for services provided the previous month	<p>Three linked tables representing episodes of care. This includes an admission table, service table, and discharge table.</p> <p><i>Admissions:</i> Demographic information, referral source, substances used, frequency of use, living situation, arrests, employment</p> <p><i>Services:</i> Record for each date of service – units of service vary by level of care. Payment responsibility for the service.</p> <p><i>Discharge:</i> Discharge reason, substances used, frequency of use, living situation, arrests, employment</p>	<p>Client level service data for people served through the ADAP funded preferred provider system.</p> <p>Note: Does not include people receiving substance abuse treatment services outside the ADAP-funded preferred provider system. This excludes individuals receiving treatment for opioid addiction through physicians, hospital based treatment, treatment received at a private practitioner social worker, mental health counselor or substance abuse counselor.</p>	Includes demographic information; collects information that allows an evaluation of change between admission and discharge.	<p>The unique ID does not identify an individual – person level data can't be directly linked to other sources of data.</p> <p>Includes only direct treatment (Outpatient, Intensive Outpatient, Residential, Hub/Methadone) services – doesn't cover the full continuum of care. Units of service provided data is unreliable due to changes in unit measures over time – for instance, reporting changed from 15 minute increments to an encounter.</p> <p>This data is in Microsoft Access</p>		<p>ADAP internal reporting</p> <p>VDH Dashboard</p>
Indicators for analysis	<ul style="list-style-type: none"> • <i>Service utilization</i> • <i>Trend analysis</i> • <i>Outcomes evaluation</i> 					

Data Source	<i>STDMIS - STD surveillance dataset</i>					
Sponsors	CDC created the data system that is locally maintained by the HIV/AIDS, STD, and Hepatitis Program Chief.					
Most current year available	Annual data available 1996 through the present - approximately 2,000 cases per year.					
Frequency	Design	Population (units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Database is updated as labs/case report forms come in, several times a week and then as case investigators interview patients. Data is uploaded to CDC at the start of each week.	Chlamydia, gonorrhea, and syphilis are reportable infections. STDMIS archives case report forms, lab results, risk profile, and treatment information for each reported case.	Any Vermont resident who is diagnosed with a reportable sexually transmitted infection (STI), regardless of state they are tested in.	All STI diagnoses among Vermonters, including treatment information and risk profile of the case and their sexual partners.	Currently the system is not set up to receive electronic reporting.		Integrated Epidemiologic Profiles for HIV/AIDS Prevention and Care Planning Summary Reports
Indicators for analysis	<ul style="list-style-type: none">• Demographic factors (age, sex, race/ethnicity, residence, sexual orientation)• Diagnostic and treatment info (earliest date, provider, facility, continuation of care)• Information on risk behaviors					

Data Source	<i>Vermont Cancer Registry</i>					
Sponsors	http://healthvermont.gov/research/cancer_registry/registry.aspx Data should be requested from the Cancer Registry Chief of the Vermont Cancer Program at the Health Department.					
Most current Year available	Complete data available 1994 through 2010.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Data is entered on a regular basis, as diagnoses occur. A diagnosis must be entered into the registry within 120 of diagnosis or admission.	This is a registry; diagnosis must be entered into the registry within 120 of diagnosis or admission.	Any Vermonter with an in situ or malignant cancer diagnosis.	All cancer diagnoses among Vermonters.	Lag in reporting by two years, no data prior to 1994, and small numbers for some cancers mean some incidence data require suppression.		Incidence Maps and Data
Indicators for analysis	<ul style="list-style-type: none"> • Demographic factors (age, sex, race/ethnicity, residence) • Primary payer • Previous cancer diagnosis • Diagnostic info (primary site, laterality, histology, behavior, grade, Diagnostic confirmation, LN ex/Pos, staging) • Treatment info (earliest date and most definitive type of each modality) 					

Data Source	Vermont Immunization Registry (IMR)					
Sponsors	http://healthvermont.gov/hc/IMR/index.aspx Data should be requested from the immunization Registry Chief of the Vermont Immunization Program at the Health Department.					
Most current Year available	Data 2000 to current is more consistently complete. Vaccination codes changed in 1996. Data from prior to 1996 and during the coding transition is more likely to be missing.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Since it is a registry, data is updated on an ongoing basis. Data is updated at slightly different time intervals depending on how a site delivers its data. A few sites can make their information available in 'Real Time'. However, most data is available anywhere from a week to 28 days after immunization delivery.	Data is collected as a registry from 3 sources: (1) data is entered by providers, (2) data is provided via monthly batch feed into the system, (3) direct transmission of real time data from an electronic medical record. Some pharmacy data is now also fed into the system and in the next couple years more pharmacies are expected to report data to the system.	All persons born in VT since 1909 have a record in the registry and any individual that has had a vaccine in a provider practice. Also included are Vermonters who have an insurance claim through a Vermont provider and who may have been vaccinated out of state.	Unified vaccination record, especially for children. The registry allows school nurses' to access immunization data directly. The registry also allows doctors to research their own patient population's rate of immunization. The IMR can also provide data at a geographic level, uptake maps, and information on underserved areas.	Data on vaccinations prior to 2000 may be incomplete. Springfield and Rutland data feeds and delivery methods are not as strong and data is sometimes incomplete. Vermonters in Windsor county that are going to New Hampshire are sometime missing vaccination information if they are not billing a Vermont insurer.	<ul style="list-style-type: none"> • Influenza Immunization • Pneumococcal immunization • Child immunization status 	NCQA assessments Immunization Goal Tracker Healthy Vermonters 2020 IISAR-Annual Report (CDC.gov)
Indicators for analysis	<ul style="list-style-type: none"> • Vaccination type and date of vaccination • Lot numbers 					

Data Source	<i>Vermont Lead Database</i>					
Sponsors	<i>The Healthy Homes Lead Poisoning Prevention Program warehouses the database and the contacts are Michael Sullivan and Andrea Haugen.</i>					
Most current Year available	~1993 to present.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Data is added to the database as information is reported by providers and laboratories in Vermont.	Any labs completing a blood lead test is required to report to the system (a 1032 database). Blood leads are sent in many formats by all laboratories and providers that performed a test. All capillary tests that are above the acceptable level, are required by statute to be followed up with a venous blood level test.	Children who have been tested for blood lead in Vermont and residents tested out of state. This includes Vermonters and those from out of state who had the blood lead test performed by a Vermont provider or laboratory.	Database includes results from all laboratories and providers that completed a lead test.	Very old DOS database.	None	Lead Legislative Report EPHT-Childhood lead poisoning
Indicators for analysis	<ul style="list-style-type: none"> • Blood Lead Levels • Child's age at test • Confirmation and re-testing rates • District office testing vs. Provider testing • Town of Residence and/or VDH district. • Among Children who have a blood lead level $\geq 10\text{mg/dL}$ and who have had a visit from the case manager: <ul style="list-style-type: none"> ○ <i>Age of property</i> ○ <i>Rental or owned property</i> 					

Data Source	Vermont Prescription Monitoring System (VPMS)					
Sponsors	<i>This is a state funded data system. Data is warehoused by an external contractor and locally managed by the Health Department VPMS analyst.</i>					
Most current Year available	2010- to present					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Data enters the warehouse database as it is collected from pharmacies at least once every seven days.	Data is entered directly by pharmacies for all dispensed controlled substances. Data is then processed by a contractor into flat files for use by the health department.	All prescriptions for controlled substances (Schedule II-IV) dispensed by Vermont licensed pharmacies.	<p>Universal database of controlled substances dispensed in Vermont. Variables on prescription, patient and provider and dispenser.</p> <p>Data is up-to-date and entered into the system as information becomes available.</p>	<p>Raw data only accessible by two analysts. Legal restrictions on what can be released.</p> <p>This registry has many users entering data with varying levels of data training.</p>	<i>May not be possible at this point in time, but it could potentially be used as a surrogate for medication reconciliation of control substances in the future.</i>	VPMS 2012 Fiscal Year Report
Indicators for analysis	<ul style="list-style-type: none"> • <i>DEA numbers of prescribers and pharmacies</i> • <i>Substance prescribed and dosage</i> • <i>Name and date of birth of recipients</i> 					

Data Source	<i>Vital Statistics</i>					
Sponsors	http://healthvermont.gov/research/records/vital_records.aspx Data collection and management is overseen by the Health Department Vital Records Office. Information should be requested from the Vital Statistics Information Manager					
Most current Year available	Final data available for 2010, preliminary data for 2011 and 2012.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Ongoing The Vermont vital statistics system monitors: <ul style="list-style-type: none"> • births • deaths • marriages and civil unions • divorces and dissolutions • fetal deaths • abortions 	<p>Births: The physician, midwife, or other birth attendant is required to complete a birth certificate within 10 days of the birth. For hospital births, the medical records staff enters the birth information into the Electronic Birth Registration System (EBRS) and submits to the Health Department. The legal birth certificate is printed at the hospital and sent to the town of birth for registration.</p> <p>Deaths: A physician is responsible for completing the death certificate, though the job is often delegated to the funeral director (with the exception of the cause of death). Information needed to complete the death certificate is obtained from the family of the deceased. Once a death record is completed in the Electronic Death Registration System (EDRS), it is registered and the death certificate is available to the town clerks for filing.</p> <p>Fetal deaths and abortions: Reports of fetal death and induced termination of pregnancy (abortion) are sent directly to the Department of Health by the physician, hospital, or clinic that performs the procedure. Reports are for statistical purposes only, are not public records, and are destroyed after five years.</p> <p>All vital records received are data entered and stored electronically, and data from births and deaths which occur in Vermont is transmitted to the National Center for Health Statistics to become part of a national database.</p>	<p>VT occurrences. Includes all resident births and deaths which occur in Vermont. Transcripts of resident birth and deaths which occur in other states are mailed to Vital Records.</p> <p>Fetal deaths and abortions are Vermont occurrences, but does not include events that occur to Vermont residents while outside Vermont (unlike birth and deaths, which are reported for VT residents by other states).</p>	<p>Births and deaths are a census of all births and deaths for Vermonters.</p> <p>There is other information that is collected as part of the birth and death certificate that we can use for analyses.</p>	<p>This is not real time and we can only look at this annually. In addition it sometimes takes a little while for each annual file to be finalized.</p>		<p>Vital Statistics Annual Bulletins</p> <p>Healthy Vermonters 2020</p>

<i>Vital Statistics (cont.)</i>	
Indicators for analysis	<p>Commonly used information from the death certificate—other information also available:</p> <ul style="list-style-type: none"> • Underlying cause of death <p>commonly used information from birth certificate—other information also available:</p> <ul style="list-style-type: none"> • Birth weight • Smoking during pregnancy • Race/ethnicity • Gestational age • Use of prenatal care <p>Mother's Weight and Height</p>

Data Source	<i>VMAP Access Database and CAREWare</i>					
Sponsors	The data system is maintained by the epidemiologist in the HIV/AIDS/STD/Hepatitis unit of the Infectious Disease Section.					
Most current year available	2000-2013					
Frequency	Design	Population (units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Database is updated as new information is reported to the health department; quarterly; and semi-annually depending on the source of the information. Data is uploaded to HRSA both quarterly and annually.	Eligibility, demographic and service provision information.	Vermont residents diagnosed with HIV/AIDS and with a FPL of 500% or less.	All HIV/AIDS service information in Vermont.		None	Integrated Epidemiologic Profiles for HIV/AIDS Prevention and Care Planning, Ryan White Services Report (RSR), ADAP (i.e. VMAP) Quarterly Report (AQR), Integrated Plan for HIV Services and Prevention.
Indicators for analysis	<ul style="list-style-type: none">• Demographic factors (age, sex, race/ethnicity, residence) for people receiving a care service listed above [strengths]• Medication Adherence info for Treatment Cascade• Service information related outpatient ambulatory, medical nutrition therapy, mental health, medical case management, dental, medication adherence)					

Claims and Discharge data

Data Source	Vermont Healthcare Claims Uniform Reporting and Evaluation System (VHCURES)					
Sponsors	GMCB/DFR http://gmcboard.vermont.gov/vhcures					
Most current Year available	Data set starts with claims incurred since Jan 1, 2007.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
The data set is continually changing and adjusting. New claims are fed in on a regular basis, which data users should be cautious of when performing analyses. Typically adjustments and payments are completed within 9 months of the date of service.	<p>Insurers (and Medicaid) covering more than 200 Vermont lives are required to report paid claims data to this database.</p> <p>BluePrint has a subset of this data set that has additional data points (<i>see VHCURES_BluePrint</i>)</p>	Individuals with paid insurance claims.	<p>We are able to see the amount paid for services.</p> <p>Information is at an individual level and we can see most encounters with the health care system (primary care, hospital, ED, pharmacy), as long as a claim is filed.</p>	<p>For some patients (self-pay) we cannot fully see their utilization of the health care system, as we can only see paid insurance claims.</p> <p>We cannot speak to individuals that are not utilizing the health care system, including those who do and do not have insurance.</p>		
Indicators for analysis	<i>We are in the process of learning more about this data set and the various claims related indicators that we can report on.</i>					

Data Source	BluePrint VHCURES data set					
Sponsors	Data maintained by external vendor (OnPoint Health Data, ME). An analytics group at Onpoint is tasked with doing a variety of analyses for the BluePrint staff using this data set.					
Most current Year available	2007 through 2012.					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Data in the VHCURES data set is updated as claims are paid and processed. The extract for BluePrint analytics is updated as appropriate (the plan is every 6 months). Eventually, real-time access through a virtual “work bench” with Onpoint has been discussed to improve access to the data.	Follows same data collection process as the broader VHCURES data set. The BluePrint version of VHCURES has additional value added including a flag for BluePrint practices. There is also additional information about attribution (to either a participating or non-participating BluePrint practice).	Paid claims of Vermont residents. (same as VHCURES with additional value added).	It is useful for measuring expenditures, and person level information among Vermonters utilizing the health care system.	Since this only includes paid claims, we do not have information on what was originally included on the submitted claims or how the claim was adjusted.		Data used for evaluation in the 2012 Annual Report
Indicators for analysis	<ul style="list-style-type: none">• Information related to monthly membership• Practice level attribution• Expenditures• Service utilization• Chronic Disease information• CRG status					

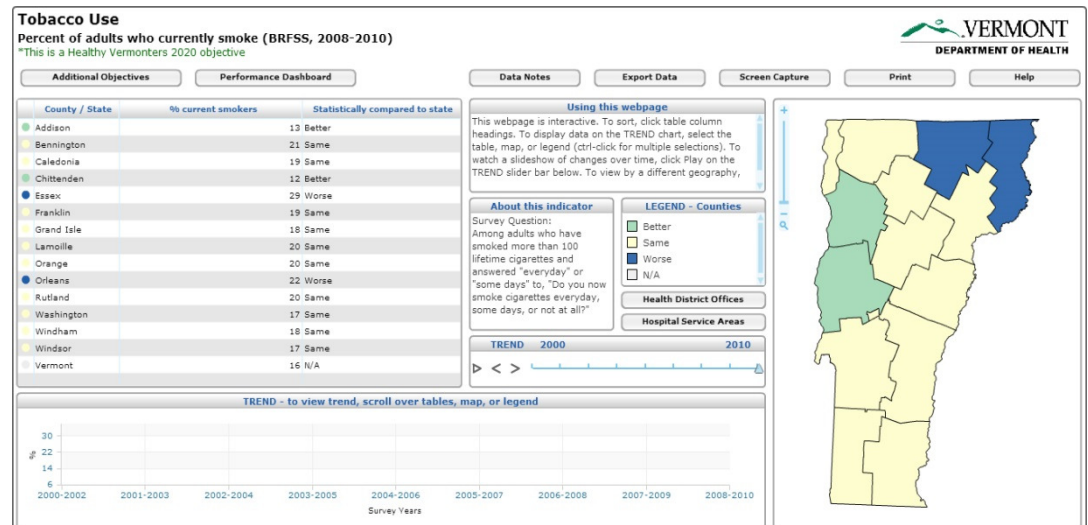
Data Source	Vermont Uniform Hospital Discharge Data Set (VUHDDS)					
Sponsors	The Health Department has an MOU with GMCB to manage and report on the hospital data. Data should be requested through VUHDDS Coordinator at the Health Department. http://healthvermont.gov/research/hospital-utilization.aspx					
Most current Year available	Chronic Disease analyst can access complete data for 2002 through 2009. Data files from 1980 through 2001 are not yet available to the Health Department users in SPSS or SAS formats, but may be upon request and approval. The earlier data years do not include as many variables or types of records. *2010 data is expected to be available soon—still waiting on data from bordering states as of 6/6/2013					
Frequency	Design	Population (Units)	Strengths	Limitations	VHCIP Measures	Reports citing data source
Annually updated (However, sometimes there are delays getting info on Vermont residents from hospitals in bordering states).	Data include all discharges submitted by Vermont hospitals to VAHHS-NSO, which then delivers the data to the Health Department as contracted with GMCB. Data for Vermont residents discharged from hospitals in NY, NH, and MA are received from those states per MOUs. Discharges from Mental Health hospitals and inpatient treatment facilities are NOT included in the hospital discharge data set. In addition to information on up to 20 diagnosis codes associated with each visit, hospitals may report primary ecode listed at discharge, patient age, sex, geographic location of origin, procedure codes, revenue codes, and total charges associated with a visit.	The unit is a hospital inpatient, outpatient, or ED discharge/visit. For most analyses we limit this to VT residents. We can also choose to limit this to VT hospitals.	Census of all Hospital and ED visits. Based on ICD-9-CM and ecodes so we can look at charges by diagnosis or procedure.	We cannot examine re-admission as there is no person-level indicator	<ul style="list-style-type: none"> • <i>Ambulatory Sensitive condition admissions (we can use VUHDDS to look at discharge rates among these age groups)</i> <ul style="list-style-type: none"> ○ Asthma and COPD admissions in older adults ○ Heart Failure 	BluePrint Big Book Annual Hospital Reports
Indicators for analysis	<ul style="list-style-type: none"> • Up to 20 diagnosis codes mentioned at discharge (ICD-9-CM)—in order listed at diagnosis there is a principal diagnosis for all inpatient/ED discharges. • CPT/procedure codes—All discharges MAY have up to 20 ICD-9-CM procedure codes. Some discharges have no procedure codes. • First ecode listed • Charges (distinguished from paid claims) • Primary Payer • Age/Sex/Hospital Service Area 					

Data Resources:

Healthy Vermonters 2020 toolkit (includes the 'Performance Dashboard' and 'Maps & Trends')				
Controller	The Health Department (Planning and Healthcare Quality unit)			
Access	http://healthvermont.gov/hv2020/index.aspx			
Years of Data Included	The performance dashboard includes the most recent population health data and information related to program performance and updated on a quarterly basis. The 'Maps & Trends' pages currently show the baseline data of the Healthy Vermonter goals, and trend data for those years preceding baseline.			
Data tool updates	Reporting Structure	Population restrictions	Strengths	Limitations
<p>The performance dashboard is updated on a quarterly basis with performance improvement measures. The surveillance measures of the dashboard are updated as new data becomes available.</p> <p>The plan for updating the Maps & Trends data set is still being developed, but is expected to be updated annually.</p>	<p>The Dashboard reports are structured in a table format with color coding, where green signifies improvement from the previous period, yellow signifies no change from the previous period, and red signifies things are moving in the wrong direction from the previous period.</p> <p>Maps, tables, and trend pages show state level and regional level data as well as some trend data. You can view a map for each year of data and a trend graph that shows the state trend and the trend in a selected region. Map, table, and trend graphs all reflect the same data.</p>	<p>Surveillance measures are calculated similarly to the corresponding Healthy People measures. This means numbers are age-adjusted and could slightly differ from crude calculations.</p> <p>More information is available in the <i>Data Notes</i> section of the Maps and Trends pages and in the <i>Turning the Curve</i> pages of the dashboard.</p>	<p>This allows public access to performance improvement data and to the Health department's Healthy Vermonter Goals.</p> <p>These pages allow for transparency of health department goals and activities.</p>	<p>At this time data is presented by measure, and all regional subgroup information is presented together on one page. It is not possible to see all of the measures for a given region in one location.</p>
Data Sources Referenced	<ul style="list-style-type: none"> • BRFSS • YRBS • Census data • Physician's Survey • Vital Statistics • Prams • ACBS • VUHDDS • WIC data • ATS • NSDUH • School Nurses' Report 			

Screen Shot examples

Healthy Vermonters Toolkit		
Statewide Population Indicators	Maps & Trends	Performance Dashboard
HV2020 Goal: <i>A Healthy Lifetime</i> 📄		
Family Planning	County District HSA	Dashboard
Maternal & Infant Health	County District HSA	Dashboard
Early Childhood Screening	County District HSA	Dashboard
School-age Health	County District HSA	Dashboard
Older Adults	County District HSA	Dashboard
HV2020 Goal: <i>Providing for Better Health</i> 📄		
Access to Health Services	County District HSA	Dashboard
Immunization & Infectious Disease	County District HSA	Dashboard
Oral Health	County District HSA	Dashboard
Mental Health	County District HSA	Dashboard
HV2020 Goal: <i>Behaviors, Environment & Health</i> 📄		
Alcohol & Other Drug Use	County District HSA	Dashboard
Tobacco Use	County District HSA	Dashboard
Nutrition & Weight	County District HSA	Dashboard
Physical Activity	County District HSA	Dashboard
Injuries	County District HSA	Dashboard
Environmental Health	County District HSA	Dashboard
HV2020 Goal: <i>Diseases & Health Conditions</i> 📄		
Heart Disease & Stroke	County District HSA	Dashboard
Cancer	County District HSA	Dashboard
Diabetes	County District HSA	Dashboard
Respiratory Disease	County District HSA	Dashboard
Arthritis & Osteoporosis	County District HSA	Dashboard
HIV, AIDS & STDs	County District HSA	Dashboard
HV2020 Goal: <i>Public Health Preparedness</i> 📄		
Public Health Preparedness	County District HSA	Dashboard



Vermont Environment Public Health Tracking Program (EPHT)				
Controller	The Health Department, Environmental Health			
Access	http://healthvermont.gov/tracking/index.aspx			
Years of Data Included	The Tracking portal includes the most recent publicly available data for each indicator; years of data included vary by indicator but are generally available for the period 2000-2010 with 2011 and 2012 data available for several datasets.			
Data Tool updates	Reporting Structure	Population restrictions	Strengths	Limitations
The Tracking portal is updated two to three times per year depending on availability of new data and software tools.	<p>Tables, charts, and maps are presented for various environmental exposures (e.g. air, drinking water) and for health outcomes that may be related to environmental exposures (e.g. cancer, childhood blood lead levels). There are a variety of indicators, some focusing on trends and some focusing on within-state geographic comparison (with error bars); additional stratification (age, sex, etc) is available where allowed by data stewards.</p> <p>Tracking participates in two CDC data calls per year submitting Vermont data not otherwise available to the federal government (e.g. birth defects, childhood lead poisoning) for display on the National portal as well as on the Vermont portal.</p>	Nationally consistent measures are calculated per CDC definition using specified population denominators. Some indicators overlap with HV2020, but case definitions and population restrictions vary meaning EPHT crude and age-adjusted rates may differ slightly from HV2020 rates.	Twenty three states plus New York City provide standardized data to the National Tracking network allowing comparison to Vermont data. Vermont-specific measures include Standardized Incidence Ratios for specific cancers, the Blue Green Algae Tracker and the Tick Tracker.	<p>100% CDC grant funded.</p> <p>EPHT is a relatively new program and has had limited opportunity to utilize Tracking data for linkage studies due to initial focus on building the portal infrastructure.</p>
Data Sources Referenced	<ul style="list-style-type: none"> • BRFSS • Vermont Cancer Registry • VUHDDS • U.S. Census • Vital Statistics 			
Index of Health and Environmental Indicators	<div> <div> <i>Health Indicators</i> <ul style="list-style-type: none"> • Asthma • Birth Defects • Cancer • Carbon Monoxide • Heart Attack • Childhood Lead Poisoning • Reproductive Health </div> <div> <i>Environmental Indicators</i> <ul style="list-style-type: none"> • Air Quality • Climate Change • Drinking Water </div> </div>			

VT EPHT (cont.)

Screen Shot Examples

Asthma

Asthma Hospitalization - Male and Female: Annual Number and Rates per 10,000 Population

Year: 2007

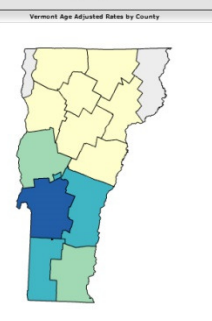
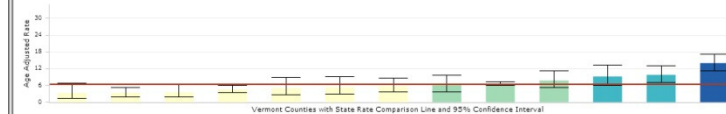
Data Notes

Screen Capture

Print

County / State	Number	Crude Rate	Age-Adjusted Rate	95% Confidence Interval (for the Age-Adjusted Rate)
ADDISON	22	6.0	6.2	(3.8 - 9.7)
BENNINGTON	33	8.9	9.2	(6.1 - 13.2)
CALEDONIA	15	4.8	5.3	(2.9 - 9.2)
CHITTENDEN	65	4.2	4.5	(3.5 - 5.9)
ESSEX	*	*	*	*
FRANKLIN	26	5.5	5.6	(3.7 - 8.5)
GRAND ISLE	*	*	*	*
LAMOILLE	8	3.4	3.2	(1.3 - 6.8)
ORANGE	12	4.1	3.5	(1.8 - 6.6)
ORLEANS	14	5.1	5.0	(2.7 - 8.9)
RUTLAND	92	14.7	13.9	(11.1 - 17.3)
WASHINGTON	20	3.4	3.3	(2.0 - 5.4)
WINDHAM	22	7.2	7.7	(5.2 - 11.3)
WINDSOR	52	9.1	9.7	(7.1 - 13.0)
VERMONT	298	6.4	6.5	(5.9 - 7.2)

*Numbers and rates based upon fewer than six cases are not displayed; when only one county has small numbers, a second county is not shown; see data notes.



Vermont Million Hearts Report

Examining the Utility of Vermont's All Payer Claims Data base

Health Statistics Chief Memo on Linking Data

TO: Heidi Klein, HS Division Director

FROM: Richard McCoy, Public Health Statistics Chief

DATE: August 13th, 2014

RE: Definition of Data Linkages

There has been some confusion with our internal and external partners on the topic of “linking datasets” or “data linkage.” The terms have routinely been misapplied and resulted in confusion on what the Health Department can provide. This memo is intended to address the confusion and provide an explanation as to what “data linkages” are and are not.

A recent example involved a state agency contacting the Health Department to ask for BRFSS data so that they could “link” it to insurer claims data (VHCURES). We stated that this was impossible since the BRFSS does not contain record-level identifiers necessary for linkage. The requesting party insisted that the linkage had to be accomplished. After further discussion, it was discovered that the requesting party wanted aggregate rates from the BRFSS for comparison to aggregate rates of VHCURES data. This “comparison” is very different from “data linkage.”

Definition

“Data linkage” is the identification of specific records (for example, an entity or person) within a dataset that refers to the records of the same entity/person in a different dataset. The linkage of two records from different sources is performed based on a common identifier (data variable) that is collected by both sources and can be reliably associated with the same entity/person. Identifiers can include, but not be limited, to:

- Name
- Date of Birth
- Date of Death
- Gender
- Mother’s Maiden Name
- Social Security Number
- Residential Address
- Telephone Number
- Place of Birth
- Medical Record Number
- Other Uniquely-assigned I.D. Number

There are a variety of software tools that statisticians use to conduct data linkage (e.g., LinkPlus). There are also different methods available to conduct matching. The primary methods are *deterministic matching* and *probabilistic matching*.

- **Deterministic data linkage / matching:** This approach uses a set of rules and any two records that meet the threshold of those rules is “determined” to be a match and can be “linked.”

For example, two datasets contain SSN, Name, Date of Birth, Sex and Zip Code. The matching rules may be established that any two records that match 100% on all five of these variables is a match and automatically linked. Alternatively, the matching rules could be that any two records that match on SSN, Name, Birth Date, and Sex (but not Zip Code) are automatically linked.

The restrictiveness of the rules will depend on the size of the population represented by the datasets and your confidence in the accuracy of the data. If the datasets represent Vermont (620,000 people), then there are less persons who share the same name and date of birth, allowing your rules to be a little less strict. If the datasets represent the entire country, then there will be many people with the same name and date of birth, so your rules will need to be much more restrictive. Additionally, if there is evidence that the dataset contains many data entry errors (transposed numbers in Date of Birth and/or SSN), then your matching rules might need to be adjusted. Alternatively, one might then wish to consider *probabilistic matching*.

- **Probabilistic data linkage / matching:** This is sometimes called “fuzzy matching.” This approach will sometimes utilize a larger set of variables than just SSN, Name, Date of Birth, etc. Each of the additional variables is assigned a weight. The weight is based on our expectation that the variable could successfully contribute to identifying a unique individual.

For example, a statistician might assign a higher weight to “Town of Birth” if the datasets (and target population) contains many towns with very small populations, and a lower weight to “Sex” due to the challenges of twins, gender changes, etc. There are two good examples that are sometimes used for probabilistic matching that can be given a high weight: Time of Birth and Birth Plurality. In Vermont, it is very, very rare that two people are born on the same date at the same exact time, and then have the same name. Therefore, if both datasets contain Time of Birth, but only one dataset contains full Name (first, middle, last) and the other contains just Last Name, the Time of Birth may be the deciding factor in whether the two records are matched.

Probabilistic matching is somewhat of an “art” – it is a balancing act to set the correct weights so that you don’t miss too many potential matches, but yet avoid matching records that are not the same entity/person. There are a variety of algorithms used to determine the sensitivity of the match. Additionally, the statistician will typically identify a “grey area” that requires manual review. In other words, you will have three sets of results – those that matched with reliability, those that did not match (high confidence they are different entities/persons), and those that are unclear. Many records can fall into this grey area and reviewing them manually can be time-consuming. Here are some examples of grey area cases that would be reviewed by a person:

Record #1	Richard H. McCoy	DOB 5/1/1970	SSN 123456789
Record #2	Rick Mccoy	DOB 1/5/1970	SSN 112345678

Probably the same person, but falls into the grey area for review due to the different DOBs, SSNs, and variation on name. These would likely be data entry issues. However, they could be different persons. (There are many “Richard McCoy’s” even with a Date of Birth in May 1970.)

Record #1	Richard McCoy	DOB 5/1/1970	Time of Birth	12:07 PM
Record #2	Henry McCoy	DOB 5/1/1970	Time of Birth	12:09 PM

These are probably *different* people. Same date of birth, similar name, but slightly different time of birth. They could also be twins. However, these may actually be the same person because the 12:09 PM might be a data entry error (and should be 12:07 PM). The different first names may be due to “Henry” actually being the middle name instead of first name.

These are relatively simple examples. There are much more complex scenarios that occur when matching records against datasets containing a million or more records, such as the Vermont Prescription Monitoring System, the Immunization Registry, and the Vital Records’ historical datasets (births/deaths).

With either deterministic or probabilistic matching, there will always be the need for manual intervention by the statistician since all of the decisions cannot be left to the software and algorithms. If two electronic medical records are incorrectly matched, it can result in adverse events or even death. Therefore, linkages must be done with careful forethought and planning, as well as appropriate time for testing, quality review, and analyses.

Once a data linkage between the records of two (or more) datasets has occurred, it is then possible to conduct analyses of greater detail about the population. It can serve as an amazingly powerful tool for public health purposes, as well as estimating service utilization and fraud prevention.

Closing Summary

Data linkage or record matching is a process by which unique records of entities/persons in separate datasets can be identified and matched to each other, thereby allowing for more comprehensive analyses of a particular population and their characteristics, behaviors and/or services. However:

- Data linkage or record matching is not the comparison of aggregate rates (such as disease prevalence) from two different datasets;
- Data linkage or record matching cannot be performed if either of the datasets does not contain variables that are considered identifiers (as listed earlier).

Also, data linkage or record matching is often restricted by federal or state law, rule or regulation. Some datasets may be considered as “confidential” and “privileged,” which significantly limits the type of data sharing and activities that can be performed without the consent of the persons within the datasets. It should never be assumed that just because two datasets contain identifiers that it would be legal and allowable to conduct linkages. Specific written agreements, such as Data Use Agreements and Memos of Understanding may be required, depending on who “owns” the data. It is always best to start with a review of the legal restrictions and consult with the department’s attorney before initiating a data linkage or sharing access to the record-level information.

If you have any questions, I’ll be happy to chat in greater detail. Thanks.

Vermont Million Hearts Report

Examining the Utility of Vermont's All Payer Claims Data base

Using Vermont's All-Payer Claims Database: A Training for Public Health Professionals

Using Vermont's All-Payer Claims Database: A Training for Public Health Professionals

Key Points from Training

- ❖ VHCURES represents most health care claims for commercially insured and Medicaid insured Vermonters; claims for Medicare insured Vermonters will be available in VHCURES in the near future.
- ❖ Reminder: health care claims data are different from health care data which are different from health data; that is, claims data are two steps removed from health data.
- ❖ There is potential for VHCURES to inform public health surveillance, research and policy questions, but, like other APCDs, VHCURES is still young in its development. Limitations related to data quality and barriers due to requirements for software, hardware, and staff expertise in using these data sets exist.
 - VHCURES can serve as a good public health resource for health care utilization and cost analyses that are cross-sectional in nature (e.g., counts of admissions or office visits). It is important to remember that this may be clearer in other states' APCDs but it is currently quite complicated in VHCURES.
 - VHCURES claims data can be good for assessing prevalence of acute diseases, not chronic diseases, which are subject to under-reporting in VHCURES. Prevalence analyses in VHCURES are limited by 1) only a few years of health care claims data are currently available in VHCURES, and 2) counts/numerators are based on whether an individual with x diagnosis/disease received health services/care for x diagnosis/disease within a given year and it is documented by a diagnosis code on a claim for health services/care. Thus, if VHCURES is considered for prevalence analyses, it is better for estimating current prevalence rather than lifetime prevalence.
 - VHCURES data may provide supplementary and/or complimentary information to existing reliable surveillance resources, such as Vermont's Utilization and Hospital Discharge Data Set, but it does not replace these resources for addressing public health questions.
- ❖ When looking to VHCURES to address a research question, consider the following:
 - Does VHCURES contain relevant data to answer the question (e.g., does not include outcome data such as lab tests; includes limited demographic data; currently Medicare data is not available).

- Who or what is in the numerator (e.g., counts of claims with primary diagnosis of asthma or any diagnosis of asthma?)
- Denominator = all members to start, can be refined by defined sub-populations from there. VHCURES includes membership data regardless of use of health care services, however dual eligibles and others with multiple sources of coverage are included in the denominator more than once.
- 9 months-1 year lag time in data being available in VHCURES
- Plan to validate your results
- Because data is continuously updated in VHCURES it is important to pay attention to the timestamp of when data was run, as this can change slightly with various versions.

Questions & Answers

What is Prevalence? The number of occurrences of a given condition/disease or health indicator during a specified time period regardless of when the condition/disease/health indicator began relative to the total population in which the condition/disease/health indicator occurs. This number includes both new cases and old cases. Although prevalence does not indicate how long a person has had a disease, it can be used to estimate the probability that a person selected at random from a population will have the disease.

How does prevalence differ from Incidence? Incidence indicates the number of new cases among those at risk for the disease/condition etc. The denominator in an incidence calculation (those at risk) varies from the total population denominator typically used in a prevalence calculation. The numerator of an incidence calculation is typically new cases, whereas a prevalence numerator captures the broader group of any of those with a disease.

What is Linked Data? Data elements that are combined from multiple sources on a specific identifier or identity (e.g., combining data on name, date of birth, zip code, etc. across multiple data sources to establish identity across data sources). Linking data offer the potential to integrate individual data across sources/systems and assess individual outcomes. An example might be linking immunization registry data to VHCURES data.

Question from program manager—If I wanted to show the prevalence of a chronic condition by the geographic area should I use BRFSS or an APCD for VHCURES?

In order to this answer we first had to clarify what was meant by prevalence of ever having a disease or the prevalence of a specific event related to a condition at one point in time. This was clarified to mean the ever having a chronic condition. The answer was then determined to be BRFSS because it captures the entire population of those with a disease regardless of insurance status.

Follow-up question from another program staffer—Couldn't you use the APCD to map burden of the disease?

Again you would need to clearly define what you mean by burden. You would also only be showing the burden among the insured population of Vermonters. Also claims data will tell you the burden within the healthcare system among users of the system. You will not be able to speak to the burden of disease outside of the clinical system. Thus, for example, the social burden of disease can be estimated from an APCD.

Vermont Million Hearts Report

Examining the Utility of Vermont's All Payer Claims Data base

Lessons Learned--VHCURES utility for chronic disease

Lessons Learned--VHCURES utility for chronic disease

The Vermont Healthcare Claims Uniform Reporting and Evaluation System (VHCURES) is a new data system for public health analysts at the Vermont Department of health. VHCURES includes only paid claims that most insurers are required to report to a state database. All private insurers that cover more than 200 Vermont lives are required to report to the database. Medicaid paid claims also appear in VHCURES and Medicare paid claims are expected to be included in the near future. Other federal insurance plans such as TRICARE do not appear in VHCURES.

At this point in time it is important to remember that VHCURES is a work in progress, the data system is improving over time and data analysts are learning about the database by using it. There is also still a lot of work to be done in validating the data, which is why many analysts may be hesitant about drawing definitive conclusions from data results.

Assuming you work for a program that focuses on chronic disease, when might VHCURES be a useful data source?

Indicator	<i>Could it be assessed in VHCURES?</i>	<i>Why could it not be assessed in VHCURES?</i>	<i>What is the ideal data source?</i>
Incidence/Point of diagnosis	VHCURES is not the best data set to analyze incidence of a chronic condition. However, it could be used for assessing the diagnosis of an acute event like cases of the flu.	Since VHCURES is claims based, we do not get information about the outcome of an encounter. While in some cases the diagnosis case could align with a condition diagnosis, this is certainly not always the case.	Registry data, more specifically registries that are aligned with condition that require mandatory reporting are the best systems for looking at population wide incidence.
Prevalence of condition	VHCURES is not the best data set to use to analyze prevalence of a chronic condition. While more investigation is needed, it could be useful for exploring the prevalence of extremely rare conditions or acute conditions, such as flu.	Since VHCURES is claimed based long term conditions, such as a chronic condition, may only appear in the list of diagnostic codes if it is a visit related to the condition, which may not happen consistently if a condition is well controlled.	At this point in time a registry or self-reported data, such as that found in BRFSS, is still the best source for prevalence of a chronic condition.
Occurrence of diagnostic or screening test	VHCURES could be used to look for the occurrence of a diagnostic test, but not the result of said test.	If the research question pertains to the result of the diagnostic test, VHCURES does not contain outcome data; it only includes billing data that would tell you if a test was performed.	Registries related to conditions that require the mandatory reporting of lab results would be the best data source. Self-reported data on the test or a test result could also be used.

Indicator	<i>Could it be assessed in VHCURES?</i>	<i>Why could it not be assessed in VHCURES?</i>	<i>What is the ideal data source?</i>
Medication Use	VHCURES could be very useful to answer questions about medication use. However, there is still a significant work to be done to get at person level data within the pharmacy claims data.	Be cautious about off label use and prescriptions that are offered at discount by large pharmacies where a claim may not be filed.	Clinical data that gets at medication prescribed and obtained would be best here. Also, self-reported information about how medication was consumed is also useful for this type of indicator.
Cost of care	VHCURES only tells you about what the insurer paid.	Cost of care in VHCURES is complicated by contracts with specific facilities that can affect how much is paid by the insurer. There is also the complication of bundled payments and capitation that we cannot decipher in VHCURES.	A charge master index that is applied to discharge data is the ideal way to look at the total cost of care as it will take into account the total cost (to the individual, insurer, and provider) for a given procedure across various service sites.